

ҚАЗАҚСТАН РЕСПУБЛИКАСЫ ҒЫЛЫМ ЖӘНЕ ЖОҒАРЫ БІЛІМ МИНИСТРЛІГІ
ҒЫЛЫМ КОМИТЕТІ
АХМЕТ БАЙТҰРСЫНҰЛЫ АТЫНДАҒЫ ТІЛ БІЛІМІ ИНСТИТУТЫ

**ҚАЗАҚ ТІЛІНІҢ
ҰЛТТЫҚ КОРПУСЫН ӘЗІРЛЕУ
ТӘЖІРИБЕСІ
(1-кезең)**

Ұжымдық ғылыми монография

Алматы, 2023

ӘОЖ 811.512.122
КБЖ 81.2 Каз
Қ 17

*Қазақстан Республикасы Ғылым және жоғары білім министрлігі
Ғылым комитеті
Ахмет Байтұрсынұлы атындағы Тіл білімі институты
«Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің
ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы»
бағдарламалық-нысаналы қаржыландыру бойынша зерттеу жобасының
(2022-2023 жж.) аясында дайындалды*

Жоба жетекшісі және жалпы редакциясын басқарған:
филология ғылымдарының кандидаты *А.Фазылжан*

Бас редактор:
филология ғылымдарының докторы *А.Жаңабекова*

Пікір жазғандар:
филология ғылымдарының докторы, профессор *Ә.Сүлейменова*
филология ғылымдарының докторы, профессор *Ш.Сейітова*

Қ 17 **Қазақ тілінің ұлттық корпусын әзірлеу тәжірибесі /** Жоба жетекшісі:
Фазылжан А. – Алматы: ЖК Асыл, 2023. – 446 б.

ISBN 978-601-7928-24-7

Ұжымдық ғылыми монографияда Қазақ тілінің ұлттық корпусы (бұдан кейін ҚТҰК) ішкорпустарын әзірлеу барысында жүргізілген теориялық зерттеулер мен практикалық ізденіс нәтижелері ұсынылып отыр. Параллель, Тарихи, Ауызша, Ономастикалық, Диалектілік және т.б. ішкорпустар құрастырудың әлемдік тәжірибесіне шолу жасалып, олардың ҚТҰК сайтындағы қолдану механизмдері көрсетіледі. Фонетика-фонологиялық, просодикалық, морфологиялық, сөзжасамдық, лексика-семантикалық, мәдени-семантикалық, метамәтіндік белгіленім әзірлемелерінің ғылыми негіздері мен алғашқы тәжірибелері сипатталады.

**ӘОЖ 811.512. 122
КБЖ 81.2 Каз**

ISBN 978-601-7928-24-7

© А.Байтұрсынұлы атындағы
Тіл білімі институты, 2023

АЛҒЫ СӨЗ

Қолдарыңыздағы ұжымдық монография 2022-2023 жылдарға арналған «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің бағдарламалық-нысаналы қаржыландыруы бойынша зерттеу жобасы аясында жүргізілген ізденістің теориялық және практикалық нәтижелеріне арналды. Теориялық нәтижелері отандық тіл ғылымы корпусытық лингвистикасы бойынша тың қорытындылар болса, практикалық нәтижесі – жалпы көлемі 40 миллионға жуық сөзқолданыстан тұратын электронды мәтіндердің репрезентатты, теңгерімді, кеңауқымды, терең аннотацияланған базасынан тұратын бірнеше ішкорпустан жасақталған Қазақ тілінің ұлттық корпусының (ҚТҰК) онлайн режимде қолжетімді, алуан түрлі параметрлер бойынша іздеуді жүзеге асыратын корпус-менеджері бар qazcorpus.kz интернет сайтының жетілген жұмысы.

Монографияда Қазақ тілінің ұлттық корпусының алғашқы кезеңі қорытындыланды, ішкорпустарды әзірлеу барысында жүргізілген теориялық зерттеулер мен практикалық нәтижелер сипаттамасы берілді. Параллель, Тарихи, Ауызша мәтіндер ішкорпусы мен Ономастикалық және Диалектілік ішкорпус құрастырудың әлемдік тәжірибесіне шолу жасалды. Фонетика-фонологиялық, просодикалық, лексика-семантикалық, сөзжасамдық, морфосемантикалық, мәдени-семантикалық белгіленім әзірлемелерінің ғылыми негіздері мен алғашқы тәжірибелері талданды. Ішкорпустар мен тілтанымдық белгіленім түрлерінің ҚТҰК сайтындағы жұмысы сипатталды.

Ұжымдық еңбек Қазақ тілінің ұлттық корпусын жасаудың концептуалды негіздеріне, ҚТҰК-ның ауызша мәтіндер ішкорпусы, ономастикалық ішкорпусы, қазақша-орысша параллель ішкорпусы, оқу ішкорпусы, диалектілік ішкорпусы, мәдени-репрезентативті мәтіндер ішкорпусы, мақал-мәтелдер ішкорпусы мен жарнама мәтіндері ішкорпусына, лингвистикалық белгіленім әзірлемелеріне арналған тараулардан тұрады. Еңбек соңынан оқырман корпусытық лингвистикаға қатысты ұғымдар мен атаулардың қысқаша сөздігін, әр тармақ бойынша қазақ, орыс, ағылшын тілдеріндегі қысқаша аңдатпаларды таба алады.

Монография отандық тіл ғылымындағы корпусытық лингвистикаға қатысты тың ғылыми тұжырымдардың жинақталуымен және корпусытық ресурс әзірлеудің алғашқы кезеңі тәжірибесінен мол ақпарат алуға болатындығымен құнды.

Еңбек авторлары, Институт ұжымы зияткер оқырман тарапынан монографияда көтерілген мәселелерге қатысты ақыл-кеңес, сын-пікірді қабылдауға әрдайым дайын. Мына байланыс жолдары арқылы хабарласа аласыздар: 050010 Алматы қаласы, Шевченко көшесі 28, Ахмет Байтұрсынұлы атындағы Тіл білімі институты, эл. пошта: tilbilimi@bk.ru

Жоба жетекшісі Анар ФАЗЫЛЖАН

1 ҚАЗАҚ ТІЛІ ҰЛТТЫҚ КОРПУСЫН ЖАСАУДЫҢ КОНЦЕПТУАЛДЫҚ НЕГІЗДЕРІ

1.1 Қазақ тілінің ұлттық корпусы: концептуалдық негіздері, бүгіні мен болашағы

1.1.1 Қазақ тілінің ұлттық корпусының концептуалдық негіздері

Қазақ тілінің ұлттық корпусы (бұдан кейін ҚТҰК) – үздіксіз толықтырылып, жетілдіріліп отыратын **ашық жоба**. Оны әзірлеуге қазақ тілін зерттейтін әлемдегі жалғыз ғылыми мекеме – Ахмет Байтұрсынұлы атындағы Тіл білімі институты ғалымдарының әлеуеті толық жеткенмен, корпус мәтіндердің ауқымды электрондық көлемін қажет етеді, тілтанушы мен IT-маманының «тандемі» негізінде жүзеге асады. Сондықтан оған тілтанушылардың ғана айналысуы жеткіліксіз, бірнеше қаржы көздерінің қолдауы аса қажет. Соған қарамастан, Институт өз қызметі шеңберінде ҚТҰК-ның мәтіндік базасын ұлғайтты, ішкорпустарын кеңейтті, метабелгіленімі мен тілтанымдық белгіленімін, сайтың, іздеу жүйесін жетілдірді. ҚТҰК – мемлекеттік тілдің лингвистикалық базасы ретіндегі инновациялық ақпараттық-анықтамалық жүйе, оның мәтіндері арнаулы белгіленім арқылы өңделіп, іздеу тетіктерімен қамтамасыз етіліп, статистикалық дәл ақпаратты барынша қысқа уақытта алуға мүмкіндік береді.

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

Тұлғалар. Қазақ тілінің ұлттық корпусын жасау мәселесін күн тәртібіне қойған Институттың бас ғылыми қызметкері, филология ғылымдарының докторы, профессор Асқар Жұбанов болды. Профессор А. Жұбанов – ұлттық тіл корпусының ғылыми негіздерін әзірлеуге зор үлес қосқан ғалым, оның қазақ тілін автоматтандырудың формалды белгілеріне, статоллингвистика мәселелеріне, математикалық және корпустық тіл біліміне арналған іргелі әрі қолданбалы зерттеулері корпустық лингвистика бойынша жүргізіліп келе жатқан ізденістің барлығы үшін тоқталмай кетпейтін бестселлерге¹, тірек еңбектерге айналды [1].

Қазақ корпусын жасау ісіне мұрындық болған ғалымдар қатарында профессор Элеонора Сүлейменованы ерекше атауға болады. Корпустық лингвистикаға

¹ Жұбанов А.Қ. Корпустық лингвистика – қазақ тіл білімінің жаңа бағыты // Тілтаным. Языкознание. ҚР БжҒМ А.Байтұрсынұлы атындағы Тіл білімі институты, сәуір-мамыр-маусым. – Алматы, 2009 жыл, №2 (34). – 3-11-б. және т.б.

қатысты терең ғылыми зерттеулерімен² бірге, ғалымның Әл-Фараби атындағы Қазақ ұлттық университетінде тұңғыш рет «Қолданбалы лингвистика» мамандығын ашып, профессор Асқар Жұбановты оқытушылыққа шақырып, осы саланың алғашқы кәсіби мамандарын әзірлеген ұстаздық еңбегі бүгінде өз жемісін беріп жатыр [2]. Ахмет Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің өзегін осы мамандар құрап, олар отандық корпустық ресурсты әзірлеуші авангард құрам болып отыр.

Қазақ тілінің ұлттық корпусын жасау ісіне Ахмет Байтұрсынұлы атындағы Тіл білімі институтының ұжымы түгел жұмылды, оның бүгінгі деңгейінің барлық жетістігі, озық тәжірибесі мұнсыз мүмкін болмас еді. Осы істе корпустық лингвистика бойынша іргелі, қолданбалы зерттеулер жүргізе жүріп, корпустық тіл білімі мамандарын даярлау, оқулық жазу ісімен қоса, корпустың мәтіндік базасын жинау, тілтанымдық белгіленімін әзірлеу, интернет сайт параметрлерін қалыптастыру, сайтқа тілдік базаны орналастыру сияқты толып жатқан күрделі ірілі-ұсақты жұмыстарды тынбай атқарып жүрген мықты маман филология ғылымдарың докторы, профессор Айман Жаңабекованың³ еңбегі орасан зор [3].

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

ҚТҰК және ұлттық бірегейлік. Қазақ тілі ірі тілдердің бірі ретінде виртуалды кеңістіктегі бірегейлік үшін жүріп жатқан әлемдік күреске тартылып

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

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

отыр. Саяси-әлеуметтік жақтан мемлекеттік тіл ретінде жоғары мәртебеге ие болғанымен, ол ақпараттық мүмкіндіктер жарысында әлі де қалып келеді. Қазақ тілінің ұлттық корпусы – мемлекеттік тілді барлық көркемдік құралдары мен мүмкіндіктерін толықтай қамти отырып сипаттайтын және оның мәртебесін арттыра түсетін заманауи тиімді құралдардың бірі ретінде ұлттық мәдениетті дамытуға, ұлттық-мемлекеттік бірегейлікті нығайтуға, ортақ мәдениетті қалыптастыруға, сол арқылы Қазақстан қоғамының ұйысуына мүмкіндік жасайды.

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

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

ҚТҰК және қазақ тілін оқыту. ҚТҰК тіл оқыту, оқулық, оқу құралдарын, түрлі типте жаттығу, онлайн-мультимедиялық сервистер әзірлеуде кең қолданыс табады, өйткені ҚТҰК оқыту мәтіні үшін аса маңызды сөздердің жиілігі мен контексті, мәтін типі мен жанры туралы ақпаратты жүйелі түрде бере алады.

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

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

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

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

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

Тенгерімділік принципі. Бұл принцип әртүрлі кезеңдегі, жанрдағы, стильдегі, автордағы және т.б. тіл фактілерінің көрініс табатын қолданыс салалары мен жағдаяттарындағы мәтіндерді корпуста пропорционалды бейнелеуді, тиісті тарихи кезеңге тән мәтіндерді сол кезеңдегі әрбір жанр, стильдің үлесіне сәйкес түрде корпусқа енгізуді білдіреді. Мәселен, ҚТҰК-да ХІХ ғасырдағы мәтіндер ішінде көркем әдебиет, ал қазіргі заманғы мәтіндер арасында публицистикалық, сонан соң ғылыми мәтіндер стильдегі үлгілері басым болуы тиіс. ҚТҰК-ның қазіргі деңгейінде бұл принцип Негізгі корпуста ғана жақсы сақталып тұр. Ал ауызша мәтіндер, тарихи мәтіндер ішкорпусында мәтіндік базаның әлі де толығыуы қажеттілігінен ішінара ғана жүзеге асып отыр. Әрине, нақты тілге мінсіз ұқсастық жасау мүмкін емес, бірақ әр уақыт кезеңі үшін корпустағы жанрлардың үлесін анықтауды шешуші түрткіжайт тілдің табиғи тіршілігіне мүмкіндігінше жақындыққа ұмтылу болып табылады.

Тілтанымдық аннотациялармен қамсыздандырылу принципі. Терең аннотациялау корпусқа тән барлық тілтанымдық белгіленіммен қамсыздандыруды білдіреді, тілтанымдық аннотацияның ең танымалы әрі ауқымдысы морфологиялық белгіленім болып табылады. Корпусқа сонымен бірге фонетикалық, просодикалық, лексика-семантикалық, фонологиялық, синтаксистік, сөзжасамдық және т.б. белгіленімдер тән. ҚТҰК белгіленім жүйесінің басқа корпустардан үлкен өзгешелігі – мәдени-семантикалық және идеографиялық (фразеологиялық) белгіленімдердің әзірленуі. Қазақ тілінің этномаркерлік лексика-фразеологиясын қамтитын мәдени-семантикалық белгіленім ҚТҰК-ны рухани және материалдық ұлттық мәдениет туралы біліммен қамтамасыз етіп, пайдаланушыларды қазақ мәдениеті негізінде бірыңғай ұлттық-мемлекеттік бірегейлікке ұйытуға жағдай жасайды. Белгіленімнің мұндай түрі ҚТҰК-ның басқа тіл корпусынан айырмашылығы мен өзіндік ерекшелігін қамтамасыз етеді, өйткені мәдени-семантикалық белгіленім қою

тілдік корпустар жасаудың әлемдік тәжірибесінде алғаш рет қазақ тілінің корпусында қолға алынып отыр. Бүгінгі корпустың синтаксистік белгіленімі ғана іске қосылмады.

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

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

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

1.1.2 ҚТҰК бірінші кезең басымдықтары туралы

Осыдан сегіз жыл бұрын жариялаған мақаламызда қазақ корпусының концептуалды параметрлері туралы ойларымыз «орынсыз көрінуі мүмкін, өйткені қазақ тілінің ұлттық корпусын әзірлеу мен жасау мәселесі теориялық та, тәжірибелік тұрғыдан да толықтай шешімін таппаған»⁴ деп көрсеткен едік [4]. Қазіргі кезде бұл ой әбден ескіргені, ақиқатқа жанаспайтыны анық. Осы уақытқа дейінгі аралықта бұл мәселе қазақ тілтанушылары шоғырланған Ахмет Байтұрсынұлы атындағы Тіл білімі институты ұжымының зор еңбегінің арқасында шешімін тапты. Алдағы уақытта көтеретін мәселелер біршама

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

жасақталған «жас» ресурсты – Қазақ тілінің ұлттық корпусы мен оның сайтын заман талабына қарай жетілдіру туралы ғана болмақ.

Тілдік корпус жасаудың жалғыз ғана басымдығы болады, ол – терең аннотациялар берілген мәтіндердің ауқымды массивтерін жасау арқылы лингвистикалық зерттеулердің сапасы мен олардың эмпирикалық базасын жақсарту екені мәлім. Дегенмен дамыған елдердің тілдік корпус жасаудағы тәжірибесі мен тілдік жағдаяттары отандық жағдайдан ерекшеленеді. Бүгінгі таңда ең ірі тілдік корпустар инфрақұрылымы және ақпараттық технологиялары дамыған елдерде жасалғаны белгілі, сонымен қатар тағы бір айта кетер жайт, бұлар – монотілдік жағдаятпен ерекшеленетін және толыққанды дамыған мемлекеттік тілі бар елдер (Жапония, Ұлыбритания, Ресей, Финляндия, және т.б.). Бұл елдермен салыстырғанда, Қазақстан көптілді жағдаятымен және өз мемлекеттік тілін дамытуға әлі көп уақыт болмағанымен ерекшеленеді. Сонымен қатар экономикамыздың қарқынды даму жетістіктерін ерекше ағай отырып, дегенмен де елімізде лингвистикалық инновациялармен байланысты, соның ішінде қазақ тілін бағдарламалауға (корпустық лингвистердің жетіспеуі, қазақ тілін жетік меңгерген компьютерлік бағдарламалар жасаушы IT-мамандардың болмауы) қатысты ақпараттық технологиялардың (ұлттық лингвистикалық корпустың инфрақұрылымы) даму деңгейі төмен екенін мойындау қажет. Осыған байланысты қазақ тілінің корпусын жасауда тілтанушылар таза ғылыми сипаттағы қажеттіліктерді, яғни зерттеушілердің корпусқа мұқтаждығын ғана емес, сонымен қатар қарапайым тұтынушының, корпусты қолданушылардың – Қазақстан Республикасы азаматтарының қажеттіліктерін де жігі ескеру қажет. Сондықтан қазақ тілінің корпусын жасаудың басымдықтарын дұрыс айқындау немқұрайлы қарайтын мәселе емес. Қазіргі таңда қазақ тілінің қолданылу аясын кеңейту және қызмет ету сапасын арттыруға қатысты туындап отырған мәселелерге байланысты тілдік жағдаятты талдай отырып, қазақ тілінің ұлттық корпусын жасаудағы **бірінші басымдық** – мемлекеттік тілдің ұлттық-мемлекеттік бірегейлігінің көрсеткіші ретіндегі рөлін орнықтыру, яғни көпмәдениетті қазақстандық қоғамды біріктіруші түрткіжайт ретіндегі қазақ тілінің базасын нығайту деп санаймыз.

Бұл басымдықты қалай жүзеге асырдық? Әзірше екі жолмен. *Біріншіден*, алғашқы кезеңде қамтылатын мәтіндік материалды саралауда көпшілікке таңымал мәтіндерге артықшылық берілді. Корпустың стильдер мен жанрларға және мәтін авторы тіліндегі әлеуметтік көрсеткіштерге тәуелсіз тілдің барлық қолданысын түгелдей алаламай қамтуы принципті мәселе, әрі солай болуы тиіс. Алайда бұл жағдайды жалпы қазақ корпусының негізгі мақсаты етіп ала отырып, алғашқы кезеңде-ақ бірден қолданушыға интернет желісінде қолжетімді болатын корпус мәтіндерінің базасын жасаудың мазмұнына мәдени құнды, оның үстіне отансүйгіштік пен елжандылық қасиеттердегі тәрбиелеу әлеуеті күшті мәтіндерге ерекше артықшылық беріліп, оларды көптеп жинау ісі жүзеге асырылды. Д.В.Сичинава, орыс тілінің корпусын жасауға белсенді қатысқан лингвист ғалым ізімен біз де мәдени сипаттағы мәтіндерге «жазушы беделі арқылы артық болатын тілдік мысалды»⁵ жатқызамыз. Ол туралы ғалымның

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

пікірі: «Лингвистиканың эстетикалық-мәдени тәсілдемеден тәуелсіздігі қаншама рет жарияланса да, жазба тіл – ол мәдениет тілі. Лингвистикалық жұмыстан, көбінесе, тілдік мысал емес, жазушы беделімен сәуленген жақсы тілдік мысал талап етіледі»; талаптың пайда болуы барлық лингвистикалық дәстүрлерге тән. Сондықтан, бір тілде мәдени сипаттағы мәтіндерді жинақтау сол тілдің мазмұндама берілген корпусын жасауда өзіндік құндылығы бар»⁶, – деп беріледі. Біз бұл пікірмен толықтай келісетінімізді ҚТҰК бірінші кезеңдегі электронды базасының көп бөлігін көркем әдебиет жанры мәтіндері құрағаны көрсетеді.

Мәдени репрезентативті корпус жасау тәжірибесі еуропа елдерінде бар. Мәселен, Венадағы Австриялық ұлттық корпус құрамында ұлттық мәдениетке ықпалын тигізген немістілді мәтіндер жинақталған болса, ҚТҰК-да бұған қоса, әлемдік корпус жасау тәжірибесінде тұңғыш рет мәдени-семантикалық белгіленім⁷ әзірленіп, іске қосылып отыр. Бұл – сөз болып отырған басымдықты жүзеге асырудың екінші жолы.

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

Соған қарамастан, мұндай белгіленім, ҚТҰК-ның, біздің анықтауымызша, бірінші болып отырған мемлекеттік, жалпыұлттық бірегейлік қалыптастыруға қызмет ету басымдығын жүзеге асыру құралы болмақ. Бүгінгі күні қазақ тілді кеңістік пен қазақстандық қоғамдастық мемлекеттік тілде сөйлеушілер санының артуымен, сөзжасаушылық, терминжасаушылық серпіліс сияқты оң үрдістермен, қоғамдық өмірдің барлық аясы мен арнасында қазақ тілі қолданысының артуымен, сапасының жақсаруымен, тілдік инновациялардың қарқынды түрде көбеюімен ерекшеленіп отырса да, қазақ сөзінің терең семантикаларында тұнып тұрған халық санасындағы «мәдени кодтардың өшуі», «рухани сананың көмескіленуі» сияқты қауіпті жағымсыз құбылыстардың күш ала бастауы байқалады. Сол себепті осы мәдени кодтарды ашудың, этно-таңбаларда жасырылған (лингвомәдениеттанымдық, бейбаламалық, лингвокогнитологиялық және т.б.) тілдік бірліктердегі семантиканы жұрт санасына сіңірудің таза ғылыми маңызы ғана жоғары емес, сонымен қатар қазақ этнотілдік қоғамы үшін оның бірегейлігін нығайту, рухани қауіпсіздігін сақтау және ұлттық мәдениетті батыстандыруға, рухсыздануына, рухани құндылықтарды құнсыздандыруға қарсы иммунитетін нығайтып, бекіту мақсатында тәжірибелік маңызы да өте жоғары болмақ. Себебі мәдени-семантикалық аннотациясы жасалған қазақ мәтіндерінің ауқымды көлемін онлайн-режимде пайдаланушыға бір сәтте ғана қолдануға бере алатын корпус осы тіл-

⁶ Сонда

⁷ «Мәдени-семантикалық белгіленім» терминін профессор Нұрғелді Уәлидің кеңесімен тұңғыш рет біз ұсынып, ғылыми айналымда қолданып отырмыз.

дік бірліктердің мазмұнында жасырылған құнды лингвомәдениеттанымдық білімдердің әлеуетті және тиімді тасымалдаушысы, насихаттаушысы, ұлттық санадағы жаңғыртушысы бола алар еді. Лингвистикалық аспектіде лингвомәдениеттанымдық анықтаманың берілуі профессор Н.Уәлидің идеялық және ғылыми жетекшілігімен шығарылған «Қазақ сөздігі» атты Үлкен біртөмдік түсіндірме сөздік пен 15 томдық «Қазақ әдеби тілінің сөздігін» құрастыру тәжірибесінде өңделіп, жүйелі түрге келтірілді, осылардың негізінде ҚТҰК-да пайдаланылатын «Лингвомәдениеттанымдық сөздік» ұсынылады.

Екінші басымдық параллель корпустарды, яғни қазақ-орыс, қазақ-ағылшын, сонымен қатар, қазақ-түрік, қазақ-өзбек, қазақ-қырғыз және республика аумағында өмір сүріп жатқан басқа да диаспора тілдеріндегі ішкорпустарды жасау. Бұл басымдықты жүзеге асыру үшін алғашқы кезеңде қазақ-орыс параллель ішкорпусының бастапқы нұсқасы әзірленді. Осы ұжымдық еңбекте екітілді корпус жасау тәжірибесі бойынша жинақталған ғылыми білімдер ұсынылды. Болашақта параллель ішкорпусты әлемдік, халықаралық, аймақаралық, туыстық және елдегі диаспора тілдерінің мәтіндерімен толықтырып, сәйкестендеру қазақ тілінің ұлтаралық қарым-қатынас тілі мәртебесін нығайтуға, ғылыми-лингвистикалық негіздемесін жасауға мүмкіндік тудыра отырып, жалпы мемлекеттік бірегейлікті нығайтуға да септігін тигізеді деп санаймыз.

1.1.3 Қазақ тілінің ұлттық корпусының бүгіні

Ұжымдық монографияда «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты зерттеу жобасының нәтижесінде анықталған ғылыми теориялық және қолданбалы білімдер жинақталып ұсынып отыр. Жоба **мақсаты** – қазақ тілінің тарихы, даму динамикасы, бүгінгі жағдайы туралы жан-жақты ақпарат беретін әрі қоғамның алуан түрлі тілтанымдық мұқтажын өтейтін, ғылыми-зерттеу және оқыту Интернет-ресурсы ретіндегі цифрланған мәтіндік базасының көлемі 40 млн сөзқолданыстан тұратын, терең аннотацияланған, теңгерімді, репрезентативті ҚТҰК әзірлеу және қызметін қамтамасыз ету болды. Осы мақсатқа жету үшін мынадай **міндеттер** орындалды:

1) ғылыми-зерттеу және оқыту Интернет-ресурсы ретіндегі ҚТҰК-ның барлық функционалды стильдерді қамтитын цифрланған мәтіндік базасының көлемі 40 млн сөзқолданысқа жеткізілді;

2) ғылыми-зерттеу жүргізудің тиімді ақпараттық-инновациялық дереккөзі болуы үшін ҚТҰК-дағы лингвистикалық белгіленім (тіл деңгейлеріне сәйкес түрлі талдау) түрі кеңейтілді: а) лексика-семантикалық, ә) мәдени-семантикалық, б) просодикалық белгіленім әзірлемелері жасалды, корпус сайтына енгізілді;

3) ҚТҰК-ға енетін (бұрын қолға алынбаған) жаңа ішкорпус түрлері: 1) «Қазақша-орысша параллель ішкорпусы»; 2) «Ауызша тіл ішкорпусы»; 3) «Тарихи мәтіндер ішкорпусы (XV-XX ғасыр)» әзірленді;

4) жаңа ішкорпустың әрқайсысына тән лингвистикалық және экстралингвистикалық (метамәтіндік) белгіленімдер әзірлемесі жасалып, сайтқа енгізілді;

5) ғылыми-зерттеу жүргізу үдерісінде пайдаланушыға мол мәлімет беру әрі жедел табу мүмкіндігін қамтамасыз ету үшін ҚТҰК сайтының «Іздеу» (Поисковик) жүйесі тәсілдері түрлендірілді, жетілдірілді;

6) корпусты қолданушыға ҚТҰК-ға енгізілген ақпараттар мен дерек-көздерді ғылыми-зерттеу жұмыстарына пайдаланудың әдіс-тәсілдері туралы анықтамалық нұсқаулық әзірленді;

7) қазақ тілін ана тілі ретінде білетін және оны жаңадан үйренуші тіл-ұстармандарға қазақ тілінің барлық қабатын оқытуға бағытталған (сөз байлығын, сөздердің мағыналық өрісін, мәдени-семантикалық және ұлттық-танымдық мазмұнын) лингводидактикалық мәлімет беретін интерфейсі жасалды, сайтта іске қосылды;

8) отандық Корпустық лингвистика ғылыми пәнінің мазмұны ҚТҰК әзірлеу барысында жинақталған теориялық және практикалық жаңа білімдер жүйесімен байыды.

Жоба алғаш рет қазақ және орыс тілінде жарық көрген шығармалар мәтіндері теңестіріліп, «Қазақ-орыс параллель ішкорпусының»; қазақ тіліндегі ауызша тілден жинақталған таспаға жазылған аудиомәтіндердің «Қазақ тілінің ауызша корпусының» аннотацияланған базасы; тарихи мұра мәтіндерінің «Қазақ тілінің тарихи мәтіндер ішкорпусының» аннотацияланған базасын жасауымен және қазақ сөздерінің синергетикасын көрсететін лингводидактикалық сипаттағы мәдени-семантикалық белгіленімдердің әзірленуімен, ҚТҰК интерфейсінің ғылыми-зерттеу және оқыту Интернет-ресурсы ретінде пайдалануға бағытталуымен айқындалады.

Теңестірілген параллель мәтіндер корпусы – ғылыми зерттеулердің, әсіресе аударма жасаудың теориясы мен практикасының аса тиімді құралы. Мәтіндердің параллель корпустары көмегімен мынадай жұмыстарды жүзеге асыруға болады:

- екітілдік және көптілдік аударма сөздіктер құрастыру;
- машиналық аударма жүйелері үшін сөздіктер құрастыру және оларды толықтыру;
- мәтіннің терминологиялық және фразеологиялық бірліктерін аудару;
- өзара теңестірілген мәтіндер корпустары мен олардың аудармаларын компьютер жадына жинақтау арқылы аудармалық жады бар жүйелер құру, нәтижесінде автоматты аударманы толық жүзеге асыру.

Параллель корпустар лексиканы автоматты классификациялауға мүмкіндік береді. Лексиканы автоматты түрде классификациялау мәтін мағынасын автоматты түрде түсінудің (нейрожүйе) ең бір тиімді әдісі⁸.

⁸ Беляева Л.Н. Лексикографический потенциал параллельного корпуса текстов // Труды международной конференции «Корпусная лингвистика – 2004». – СПб., 2004. – С. 55-64.; Митрофанова О.А., Грачкова М.А., Шиморина А.С. Автоматическая классификация лексики в параллельных текстах (на материале текстов из Русско-словацкого корпуса параллельных текстов PARUS) // Материалы V Международной научно-практической конференции «Прикладная лингвистика в науке и образовании: лингвистические технологии и инновационная образовательная среда». – СПб.: «ЛЕМА», 2010. – С. 231-235.; Гарабик Р., Захаров В.П. Параллельный русско-словацкий корпус // Труды международной конференции «Корпусная лингвистика – 2006». – СПб., 2006. – С. 81-87.

Бұрын-соңды жасалмаған ауызша тіл мәтіндерінің ішкорпусы әзірленді. Бірақ оның мәгіндік қоры әзірше шағын. Әлемдік корпустарда да ауызша тіл корпустарының көлемі жазба тіл корпустарынан әлдеқайда аз. 1960 жылы Лондон университетінің колледжінде Р. Квирктың ғылыми жетекшілігімен «Обзор употребления английского языка» жобасы аясында алғашқы ауызша тіл корпусы болып табылатын ағылшын тілінің Лондон-Лунд (The London-Lund Corpus) корпусы құрастырылған⁹. Корпус көлемі – 1 млн. сөзқолданыс. Оған ауызша тілдің мәтіні ретінде радиохабар жазбалары, мәжілістер және бейресми әңгіме жазбалары алынған. Лондон-Лунд корпусы ағылшын сөйлеу тілін компьютер көмегімен зерттеу құралы ретінде ұзақ уақыт бойы аса маңызды дерекнама болып келді.

Қазақ тілінде лингвистикалық ғылыми зерттеулер негізінен жазба тіл мәтіндері бойынша жүргізілді, алайда жазба мәтіндерден ауызша коммуникацияда ғана байқалатын дискурстық жағдаяттар мен прагматикалық дйттемдерді айқындау қиын. Осындай қазақ тілінің ауызша тіл мәтіндерінің корпусын құрастыру бүгінгі күні ауызша тілді зерттеудің таптырмас құралы ретінде аса қажеттілік тудырып отыр.

Әлемдік тіл білімі, оның ішінде қазақ тіл білімі бойынша ғылыми зерттеулердің дені көбінесе жазба тілдің материалдары бойынша жүргізіледі де, ауызша тіл дереккөздері бойынша жүргізілген зерттеулер некен-саяқ. Шетел және орыс тіл білімінде Ұлттық корпустарды құрастырушылар өз мақалаларында ауызша және жазба тілде жеткізілген ойдың айырмашылықтарын көрсете отырып, ауызша тіл мәтіндерінің корпусын құрастырудың қажеттілігін атап көрсетеді¹⁰. Әдетте жазба мәтіндерде байқала бермейтін сөйлеу жағдаяты, коммуниканттар прагматикасы қарым-қатынас кезіндегі ауызша коммуникацияда айқын көрініс табады¹¹. Осымен байланысты ауызша коммуникацияны зерттеудің мәні зор. Бұндай зерттеулер жүргізудің ең оңай әрі тиімді жолы – қазақ тілінің ауызша тіл мәтіндерінің корпусын құрастыру. «Қазақ тілінің ауызша тіл ішкорпусы» жасалатын болса, ғылыми зерттеулерде тілдік қатынастағы дискурстық жағдаяттар, прагматикалық дйттемдер мәселелері кеңінен ашылып, «Сөйлеу коммуникациясы теориясы», «Тілдік қатынас», «Прагматика»,

⁹ Svartvik J., Quirk R. A corpus of English Conversation. – Lund: Gleerup, 1980.

¹⁰ Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: Учеб. пособие. – М.: Издательский центр «Академия», 2004. – 208 с.; Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005.; Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Едиториал УРСС, 2003. – С. 114.; Национальный корпус русского языка. – <http://ruscorpora.ru>; Шерстинова Т.Ю. «Один речевой день» на временной шкале: о перспективах исследования динамических процессов на материале звукового корпуса // Филология. Востоковедение. Журналистика. Серия 9. – СПб., 2009.

¹¹ Баранов А.Н., Плулган В.А., Рахилина Е.В. Путеводитель по дискурсивным словам русского языка. – М., 1993.; Николаева Ю.В. Кинетические признаки структуры устного нарратива (корпусное исследование) // Проблемы компьютерной лингвистики: Сборник научных трудов / Под ред. А.А. Кретьова. – Вып. 4. – Воронеж, 2010. – С. 193-200.; Маркасова Е.В. Риторическая энантиосемия в корпусе русского языка повседневного общения «Один речевой день» // Компьютерная лингвистика и интеллектуальные технологии. – Выпуск 7 (14). По материалам ежегодной международной конференции «Диалог» (2008) / Гл. ред. А. Е. Кибрик. – М., 2008. – С. 352-355.; Крейдлин Г.Е. Голос и тон в языке и речи // Язык о языке / Отв. ред. Н.Д. Арутюнова. М.: Языки русской культуры, 2000. – С. 453-501.

«Дискурсивтік лингвистика» сияқты тіл білімі салаларының тереңдей зерттелуіне септігін тигізеді.

Ауызша тіл мәтіндерінің корпусы үшін фонетикалық, просодикалық белгіленімдер әзірленіп енгізілді. Просодикалық белгіленімдер тілдің дыбыстық жүйесінің айтылымдағы репрезентациясын, сөйлеу орфоэпиясын, дыбыстар мен сөздер тіркесімін (дистрибуциясын) зерттеуге ықпал етеді. Ауызша тіл мәтіндері бойынша жиілік сөздіктер алуға болады. Ауызша тіл мәтіндерінің жиілік сөздігі қазақ тілін үйренушілерге арналған лексикалық минимумдер, оқулықтар құрастыруда таптырмас құрал болып табылады. Себебі ауызша сөйлеу этностық тегіне және тілдік білімді меңгеру деңгейіне қарамастан, инофонның тілді тиімді меңгеруін қамтамасыз ететін табиғи, қолайлы байланыс ортасын жасайды.

Қазақ тілінің тарихи мәтіндері (ескерткіштер) бұрын-соңды корпус базасына жинақталмаған. Ұлттық корпустар – тілдің қазіргі (синхронды) жағдайы туралы ақпарат беретін жүйе ғана емес, сонымен қатар тілде болған көне жазулардан (ескерткіштер) бастап, барлық жазу түрінде (төте жазу, 1929-1938 жылғы латын) жарық көрген мәтіндерді де жинақтайтын, тілдің өткенінен де ақпарат беретін құрал. Тарихи корпустар ҚТҰК-ны пайдаланушыға көне жазба ескерткіштерді, латын, төте жазу графикасында жазылған мұрағаттарды, тарихи мәтіндерге қатысты қажетті ақпараттарды корпус базасынан оңай және тез тауып алуына мүмкіндік береді. Тарихи корпус көне жазба ескерткіштерге ғылыми-текстологиялық зерттеулер жүргізудің таптырмас дереккөзі болып табылады.

Тарихи корпус көне жазба ескерткіштер мен төте жазуда, 1929-1938 жылдарғы латын қарпінде жазылғандықтан, мәтіндерде таныс емес сөздер, көне сөздер, араб, латын графикасында жазылған сөздер бойынша арнайы лингвистикалық белгіленімдер қоюды қажет етеді. Қазіргі кирил қарпіндегі сөз бен төте жазудағы сөзді немесе латындағы сөзді сәйкестендіретін, сондай-ақ көне сөздердің мағыналарын түсіндіретін сөздік-белгіленімдер қойылатын болады. Бұндай белгіленімдер ҚТҰК-ны пайдаланушыларға тарихи мәтіндерді оқып түсінуге, ғылыми-зерттеулер үшін қажетті ақпараттар алуға мүмкіндік береді.

Тарихи корпустар тарихи ескерткіштерді зерттеуге арналады. Оған орыс тілінің XVI-XVII ғғ. агиографиялық мәтіндердің Санкт-Петербург Корпусын (АМСК) жатқызуға болады¹². Соңғы уақытта қазақ тілінде жазба ескерткіштерді зерттеу мәселелесі біршама тоқырай бастады. Тіл тарихын зерттеуші мамандар дайындау ісі де жолға қойылмай отыр. Қазіргі таңда көне түркі жазба ескерткіштерін ғана емес, төте жазумен, латын қарпімен жазылған XX ғ. басында жазылған рухани мұрағаттарымызды сақтап қалуымыз қажет. Осы орайда ҚТҰК құрамына енетін көне түркі жазба ескерткіштерден бастап, араб, латын қарпімен жазылған жазба мұраларымызды электронды нұсқада сақтап отыратын «Қазақ тілінің тарихи ішкорпусы (XV-XX ғасыр)» алғаш рет әзірленді. Бастапқы кезеңде оған орта ғасыр жазба ескерткіштері енді.

Бағдарлама нәтижесі қоғамға қажетті әрі мемлекеттік тілдің қазіргі кезде және қолданылып келе жатқан әр кезеңдердегі қызмет етуінің барлық қырлары

¹² Азарова И.В., Алексеева К.Л., Захарова Л.А. Разметка текстовых фрагментов в корпусе агиографических текстов СКАТ // Труды международной конференции «Корпусная лингвистика – 2006». – СПб: изд-во С.-Петербург. ун-та, Изд-во РХГА, 2006. – С. 16-24.

жөнінде толыққанды түсінік беруге қабілетті, қазақ тілін зерттеу мен еліміздің әртүрлі санаттағы азаматтарының қазақ тілін үйренуіне жаңа мүмкіндіктер ашады. Сондай-ақ ұлттық және мемлекеттік тіл ретіндегі қазақ тілін кодификациялаудың, стандарттау мен жаңғыртудың көзі қызметін атқаруға, оның даму үрдістерін анықтауға қабілетті инновациялық көпқызметті электронды өнім ретіндегі ҚТҰК Интернет-ресурсы¹³ әзірленді.

Жобаның ғылыми бастамын Институтта әзірленген қолданбалы бағыттағы төмендегідей зерттеулер мен өнімдер құрады:

– компьютер жадына енгізілген Ахмет Байтұрсынұлы атындағы Тіл білімі институтының 5 миллиондық картотекалық қоры;

– терминологиялық түсіндірме сөздіктер мен екітілді терминологиялық сөздіктер;

– қазақ тілінің біртөмдік, онтомдық, екітомдық сөздіктері, онбестөмдік «Қазақ әдеби тілінің сөздігі»;

– «Қазақша-орысша сөздік» (2010), «Орысша-қазақша сөздік» (2005), «Орысша-қазақша қоғамдық ғылымдар терминдерінің сөздігі» (2010);

– «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» (2016), «Қазақ сөзформа құрылымының жиілік сөздігі» (2017), «Қазақ графемаларының жиілік сөздігі» (2018) «Қазақ ауызша тіл мәтіндерінің жиілік сөздігін» (2020) жасауға алынған аудиомәтіндер;

– «Ескі қазақ тілінің сөздігі» (2020); «Кирилл және латын графикаларында кірме араб-парсы сөздерінің түсіндірме сөздігі» (2020), «Қазақ тіліндегі араб, парсы сөздерінің түсіндірме сөздігі» (2017);

– «Қазақ әдеби тілінің аннотацияланған ұлттық корпусы» (2012-2014), «Қазақ тілінің ұлттық корпусындағы метамәтіндік белгіленімдер ұстанымдары мен әзірлемесі» (2015-2017) атты зерттеулер¹⁴.

¹³ Перцов Н.В. О роли корпусов в лингвистических исследованиях // Труды международной конференции «Корпусная лингвистика – 2006». – СПб.: Изд-во С.-Петербур. ун-та; Изд-во РХГА, 2006. – С. 318-331.

¹⁴ Сулейменова Э.Д. Государственный язык: статусное и корпусное планирование // Мир языка и деловое общение: Матер. межд. науч.-практ. конф. Алматы, 2002. С. 124-131.; Жубанов А.К., Фазылжанова А. О концепции по созданию национального корпуса казахского языка (НККЯ) // Материалы Международной научно-практической конференции «Контрастивные исследования и прикладная лингвистика». – Минск, 2014.; Жубанов А.К. Мәтіндер корпусы – лингвистер үшін аса күшті ақпараттық ресурс // «Фонетика ғылымдарының өзекті мәселелері: даму көздері мен болашағы» атты халықаралық ғылыми-теориялық конф. материалдары. – Алматы, 2012. – 186-193-бб.; Жубанов А. Тіл мен мәдениеттің лексика-фразеологиялық жүйесін қазақ тілінің Ұлттық корпусы негізінде кешенді зерттеу. / «Қазақ лексикографиясы: тарихы, тәжірибесі, болашағы» атты халықаралық ғылыми-теориялық конференция. – Алматы: «Елтаным баспасы», 2015. – 177-181-бб.; Жубанов А. Этномәдениеттік семантика белгіленімі «Қазақ тілінің ұлттық корпусында» ескерілуі қажет // «Академик Ө.Айтбайұлы және мемлекеттік тіл мәселелері» атты халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 2016 ж., 22 сәуір. – 344-348-бб.; Жубанов А. Қазақ тілінің ұлттық корпусы – қуатты ақпараттық ресурс // Актуальные проблемы диалектологии языков народов России. XVI Всероссийская конференция. – Уфа, 1-4 июня 2016 г. – 157-161-бб.; Жаңабекова А., Пірманова К. Технология программы полуавтоматической метаразметки казахского национального корпуса // Актуальные проблемы современной лингвистики и гуманитарных наук: сборник статей XI Всероссийской научно-методической конференции с международным участием. Москва, 15 марта 2019 г. – Москва: РУДН, 2019. – 727 с.;

– «Корпустық лингвистика» (2016) оқу құралы¹⁵; «Қазақ ауызша тіл мәтіндерінің жиілік сөздігі» (2020)¹⁶.

Аталған ғылыми зерттеулердің нәтижелері, сондай-ақ «Қазақ тілінің ұлттық корпусы» сайты (<http://qazcorpus.kz/>), «Қазақ тілінің әмбебап электронды сөздігі» (<https://kitap.kz/>), «Сөзтексер» электронды бағдарламасы» (<https://soztekser.kz/>) сияқты IT-ресурстар осы жобаның қолданбалы аспектідегі ғылыми бастамы болды.

Аталған жаңа ішкорпус түрлерінің әрқайсысына тән лингвистикалық (лексика-семантикалық, мәдени-семантикалық, просодикалық) және экстралингвистикалық (метамәтіндік) белгіленімдер әзірлемесі жасалып, ҚТҰК бағдарламасының енгізілуі пайдаланушыларға тілдің лексика-семантикалық, мәдени-семантикалық, просодикалық жүйесіне қатысты мол ақпарат алуына мүмкіндік берді. Лексика-семантикалық белгіленімдер әдіскерлер мен оқулық авторларына, лексикалық минимум құрастырушыларға белгілі бір тақырыптық, семантикалық топ аясындағы сөздерді жекелей немесе топтап, олардың мәтіндік контентін қоса, бір жерден тауып алуына, ал мәдени-семантикалық белгіленім қазақ тілін ана тілі ретінде білетін және екінші тіл ретінде үйренушілерге, қазақстандық жастарға ұлттық коды бар қазақ сөздерінің мәдени-уәждемелік семантикасын игеруге ықпал етеді. Бұл келешекте қазақ халқының ұлттық құндылықтарын, салт-дәстүрін білетін рухани дүниесі бай тілдік тұлғаның қалыптасуына мұрындық болмақ.

Ғылыми жоба міндеттерін шешу барысында төмендегідей әртүрлі **әдіс-тәсілдер** қолданылады: ҚТҰК құрамына енгізілетін ішкорпустардың мәтіндер базасын жасауда *жинақтау* әдісі; лексика-семантикалық белгіленім әзірлемесін жасауда қазақ сөздерін *топтау*, мәдени-семантикалық белгіленім әзірлемесін жасауда мәдени мазмұны бар сөздерді *жинақтау, дефиниция беру* (түсініктеме), мәдени бірліктердің шығуына негіз болған уәждемелік талдау әдісі; Қазақ-орыс параллель ішкорпусын әзірлеуде екі тілдегі мәтін аудармасын теңестіру (выравнивание) әдісі; Тарихи ішкорпус мәтіндері бойынша белгіленімдер енгізуде тарихи-салыстырмалы әдіс қолданылады. Ауызша ішкорпусты құрастыруда ауызша мәтіндерді *таспаға жазу*, аудиомәтіндерді жазбаша *электронды пішінге келтіру*; просодикалық белгіленім әзірлемін жасауда *анализ және синтездеу, эксперименттік* әдістер; ҚТҰК ішкорпустары бойынша аталған лингвистикалық белгіленімдер әзірлемесін жасауда және пайдаланушыға қажетті лингвистикалық және экстралингвистикалық ақпаратты жылдам тауып алуға мүмкіндік беретін «Іздеу» жүйесі интерфейсін жетілдіруде *компьютерлік-бағдарламалық* әдістер; отандық Корпустық лингвистика ғылыми пәнінің мазмұнын ҚТҰК әзірлеу барысында жинақталған теориялық және практикалық жаңа білімдер жүйесімен толықтыруда *сипаттамалы* әдістер қолданылды.

Лексика-семантикалық белгіленім қазақ тіліндегі барлық сөзді тақырыптық және лексика-семантикалық топтарға жіктеп, топтастыру арқылы

¹⁵ Жұбанов А., Жаңабекова А. Корпустық лингвистика. – Алматы, 2016. – 336 б.;

¹⁶ Жұбанов А., Жаңабекова А., Токмырзаев Д. Қазақ ауызша тіл мәтіндерінің жиілік сөздігі. – Алматы, 2020.

жүзеге асырылды. Бұл жіктеме корпус үшін ғана емес, тіл білімінің лексикология, семантика салалары бойынша да ірі теориялық зерттеулердің туындауына жол ашады. Сонымен қатар корпуста енгізілген лексика-семантикалық белгіленімдер арқылы пайдаланушылар өзіне қажетті лексика-семантикалық топтағы сөздерді оңай және тез тауып алып, ҚТҰК-дағы мәтіндік базадан мол материал (мысалдар) ала алады.

ҚТҰК-ға енгізу көзделген **мәдени-семантикалық белгіленімдер** қазақ тіліндегі ұлттық коды бар, мәдени семантикаға ие сөздердің шығу тарихы, уәжі, сөздің мағыналық құрылымындағы халықтың дүниетанымы, мәдени өмірі, ұлттық болмысы туралы ақпарат береді. Бұл белгіленім қазіргі кітап оқу дағдысы тарыла бастаған ақпараттық қоғамда өмір сүріп жатқан жастардың, ҚТҰК сияқты тиімді цифрлық құралды пайдалану арқылы, қазақ тіліндегі сөздердің ұлттық кодын, мәдени семантикасын меңгеруіне, сөз құрылымындағы ұлттық құндылықтарымызды саналарына сіңіре түсуіне ықпал етеді.

Просодикалық белгіленімдер әзірлемесін жасау осы жобада жоспарланып отырған ауызша тіл корпусы мәтіндеріне (аудиожазба) қазақ тілінің дыбыстық-просодикалық ерекшеліктері бойынша мол ақпарат беріп, бұл келешек тіліміздің орфоэпиялық және орфографиялық жүйесін дұрыс тануға, сөздерді дұрыс айту мен дұрыс жазуға септігін тигізеді.

Ұсынылып отырған ішкорпустардың әрбірі бойынша метамәтіндік белгіленім әзірлемесін жасау ҚТҰК-ны қолданушыларға бұл мәтіндердің авторы, аудармашысы, мәтіннің аударылған мерзімі, аудиожазбаның дереккөзі, уақыты, тарихи мәтіннің жазылған уақыты, типі, стилі, жазу қарпі және т.б. туралы мәліметтер алуға және осы белгіленімдер арқылы қажетті ақпаратты іздеп, оны әрбір ішкорпуста табуға мүмкіндік береді.

Іздеу жүйесінде мәтін стилі, мәтін авторы, мәтін жылы және т.б. параметрлермен іздеу пайдаланушыға белгілі бір стиль бойынша зерттеу жүргізуге, қандай да бір автордың шығармашылығын, тілтанымдық мұрасын зерттеуге, сөздіктер иллюстрациясы үшін мысалдар табуға септігін тигізеді. Ұсынылып отырған ішкорпустарда да метамәтіндік параметрлер бойынша іздеу жасау көзделген. Сонымен қатар лингвистикалық белгіленімдер (лексика-семантикалық, мәдени-семантикалық, просодикалық) бойынша іздеу жүйесі толықтырылып, бұл пайдаланушыға осы параметрлер бойынша қажетті ақпарат алуына мүмкіндік тудырады.

Сонымен, қорыта айтқанда, бүгінгі Қазақ тілінің ұлттық корпусында 40 миллиондай сөзқолданыстан тұратын «ақылды» мәтіндік база бар, оның 31 миллионнан астамы Негізгі корпуста, 2 миллионнан астамы Параллель мәтіндер ішкорпусында, 1 миллионнан астамы Тарихи мәтіндер, тағы 1 миллионнан астамы Ауызша мәтіндер, 3,5 миллиондайы Диалектілік ішкорпуста, 140 мыңға жуығы Мақал-мәтелдер, 130 мыңдайы Ахмет Байтұрсынұлының арнайы ішкорпусына енгізіліп отыр.

Корпустың мәтіндік базасын жасақтау үшін Тіл білімі институтының қорындағы 7000 лингвистикалық еңбектен тұратын кітапхананың электрондық нұсқасы жасалып, сайтқа (<http://tbikitap.kz/>) орналастырылды. Осы PDF нұсқадағы мәтіндерді тану үшін Институт мәтінді автоматты тану бағдар-

ламасының жобасын жүзеге асырып жатыр. Болашақта электрондық pdf нұсқадағы мәтіндер техникалық өңдеуді аз қажет ететіндей түрде корпус базасына енгізілетін болады.

Екінші, бағдарламада көрсетілген міндеттерден бөлек тағы да орындалған жұмыс – қазақ тілінің ұлттық қоры болып саналатын 4,5 миллионнан астам картотекалық базаның электронды пішінге келтіріліп, сайтқа орналастырылуы (<http://tbikartoteka.kz>)

Қазақ мәтінін қол жазудан автоматты тану бағдарламысын әзірлеу – Институт алдында тұрған болашақ міндеті. Бұл орындалғаннан кейін 4,5 миллионнан астам қазақ сөзінің контекстін мәтін түрінде корпус базасына енгізуді көздеп отырмыз.

1.1.4 Қазақ тілінің ұлттық корпусының болашағы

Әлемдік корпустық лингвистика жетістіктері мен ұлттық корпустарды жасау мен пайдаланудың озық тәжірибелеріне (Ресей Федерациясы, Финляндия, Ұлыбритания, АҚШ, Жапония және т.б.) сүйене отырып, әртүрлі лингвистикалық белгіленімдер мен метабелгіленімдер енгізілген, қазақ тілінің стильдік жанрларын толық қамтитын Қазақ тілінің ұлттық корпусының кең ауқымды, үйлестірілген әрі жан-жақты болуын қамтамасыз ету жұмыстары болашақта жүргізіле бермек. Бұл ретте ҚТҰК-ға тіл дамуының көне кезеңдерінен бастап, қазірге дейінгі әртүрлі стиль, жанр, аумақтық және әлеуметтік нұсқаларды толық қамтитын жазба және ауызша мәтіндерді, соның ішінде көркем әдебиет, публицистика, шаршысөз, сахна тілі, дайындықсыз сөйлеудің таспаға түсірілген мәтіндерін, ісқағаздарын, құжаттар, жарнамалар, медиаресурстар, интернетресурстардағы, чат-коммуникациялардағы және т.б. мәтіндерді енгізу ісін үзбей жалғастырып отыру аса маңызды.

Әртүрлі мақсатта қолдануға арналған ашық жүйе түріндегі ақпараттық-анықтамалық құрал болып табылатын ҚТҰК туралы жалпы мәселелерді жинақтай келе, оның концептуалды параметрлерін алғашқы тараушада көрсетілгендей, сақтай әрі үнемі жақсартып отырып, көлемі мен негізгі ерекшеліктерін жетілдіру бағыттары төмендегіше анықталды:

– Корпусқа енгізілетін мәтін көлемін екінші кезеңде 300 миллион сөз-қолданысқа жеткізу керек (ағылшын тілінің Оксфорд корпусында 6 миллиард сөзқолданыс, ал Орыс тілінің ұлттық корпусында 1 миллиардтан астам сөзқолданыс қамтылған). Әртүрлі графикалық жүйеде жазылған корпусқа енгізілетін мәтіндерді алдын ала филологиялық сараптамадан өткізіп, оның тізімін ұлғайтып отыру қажет.

– Қазіргі қазақ тілі үш түрлі графикалық жүйеге негізделеді: кирил (Қазақстанда, Моңғолияда, Ресей Федерациясында, Қырғызстанда, Түркіменстанда, Өзбекстанда және т.б. бұрынғы Кеңес өкіметінің басқа да елдерінде), латын (Түркия, Германия, АҚШ және т.б. батыс елдерінде), араб, яғни төте жазу (Қытайда, Ауғанстанда, Пәкістанда және т.б. Таяу Шығыс елдерінде). Әртүрлі графикалық жүйеде жазылуына қарамастан, қазақ тілінің сөз байлығын көрсететін барлық мәтіндерді электронды базаға мүмкіндігінше түгелдей жи-

нау үшін және Қазақ тілінің ұлттық корпусын зерттеушілер үшін ғана емес, Қазақстандағы, басқа елдердегі барлық пайдаланушыларға қолжетімді болуы үшін әртүрлі графикалық жүйедегі мәтіндердің кең ауқымын корпуста енгізіп, үш графикалық жүйеде жазылған, лингвистикалық белгіленімдер қойылған электрондық нұсқадағы мәтіндердің көлемін, олардың әрқайсысының Корпус құрамындағы арасалмағын шынайы өмірдегі қолданысының көлеміне, қызметтерінің алатын орнына қарай белгілеп алу керек.

– ҚТҰК тарихи және қазіргі мәтіндердің толық жиынтығы ретінде қазақ тілінің даму тарихы туралы жалпы және оның кейбір жекелеген мәселелері туралы өте көлемді ақпарат беруі тиіс, сонымен бірге үздіксіз үдерістегі көзге байқала бермейтін тілдік өзгерістерді сипаттай алуға септігін тигізе алуы керек, сондықтан оның диахрониялық ауқымдылығын кеңейту, яғни тарихи мәтіндер көлемін арттыру үшін көне жазба мәтіндерден бастап, Қазақстан тарихындағы әр кезеңде және әртүрлі графикалық жүйеде жазылған барлық қолжетімді мәтіндерді корпуста енгізу міндеті алда тұр. Электрондық нұсқаға түсірілген, лингвистикалық белгіленімдер қойылған үш графикалық жүйедегі тарихи мәтіндердің көлемі мен арасалмағын хронологиялық пропорция бойынша белгілеу абзал.

– ҚТҰК үйлесімділігін, жан-жақтылығын, ауқымдылығын қамтамасыз ету үшін корпуста қазақ тілінің барлық жанры мен әртүрлі стильде жазылған мәтіндерін енгізуді тоқтатпау қажет: көркем проза, поэзия, ауыз әдебиеті үлгілері, драматургия, мемуарлық-өмірбаяндық әдебиет, кітаби публицистика, жаңалықтар, ғылыми, ғылыми-техникалық және ғылыми-көпшілікке арналған мәтіндер, оқулық мәтіндері, діни және діни-философиялық мәтіндер, өндірістік-техникалық мәтіндер, іскери мәтіндер, заң мәтіндері, тұрмыстық (сөйлеу тілі) мәтіндер, эпистолярлық мәтіндер (жеке хаттар, күнделіктер т.б. мәтіндері), электрондық нұсқадағы ауызша мәтіндер, медиаресурстар мен интернет ресурстар, әлеуметтік желілер мен чаттар және т.б. Қазақ тілінің синхрондық және диахрондық қолданысындағы пропорциялары ескеріле отырып, әрбір жанрдың, стильдің, жазба және ауызша формадағы мәтін көлемі анықталатын болады.

– Болашақта омонимдері ажыратылған өзге тілділерді оқытуға негізделген Оқу ішкорпусы; сәйкестендірілген қазақша-ағылшынша; қазақша-түрікше және т.с. екітілді Параллель ішкорпустар; Мәдени репрезентативті мәтіндер ішкорпусы, Поэтикалық мәтіндер ішкорпусы; Дайындықты және Дайындықсыз сөйлеу ішкорпустары; Мультимедиялық ішкорпус, Бала тілі ішкорпусы, «Қателер» (ортологиялық) ішкорпусы, Тұрмыстық тіл «диктофонмен бір күн» ішкорпусы, Әлеуметтік желі тілі ішкорпусы, Ғылым тілі ішкорпусы іске қосылуы тиіс.

– Әр тілдегі корпус құрастырудың озық үлгілеріне сүйене отырып, ҚТҰК мета белгіленім мен лингвистикалық белгіленімдері мынадай параметрлерге толық сәйкес болуы қамтамасыз етіледі: а) мета белгіленім және лингвистикалық белгіленімдер – корпуста енгізілген мәтіндерге қойылатын тұтастай немесе оның жекелеген бөлігі туралы қосымша экстралингвистикалық және лингвистикалық, социолингвистикалық, әдебиеттанымдық, библиографиялық,

және т.б. ақпараттар жиынтығы; ә) метабелгіленім мен лингвистикалық белгіленімдер тіл қолданысы мен өзгерісі туралы ақпарат беретін шынайы дереккөз болып табылады; б) корпустардағы белгіленімдер дәл және жан-жақты берілуіне қарай олардың қазақ тілін оқытудағы, сондай-ақ әлеуметтік, ғылыми құндылығы да арта түседі; в) метабелгіленімдер мен лингвистикалық белгіленімдер қою бағдарламасын үнемі жетілдіріп отыру қажет; г) мәтіндерге белгіленімдер қою жұмысы арнаулы бағдарламалар мен толықтырылып отыратын сөздіктер көмегімен автоматты және жартылай автоматты түрде кезең-кезеңімен жүзеге асырылады.

ҚТҰК көлемі мен концептуалды параметрлерін негізге ала отырып, тілтанушылар мен бағдарламашылар бірігіп, оны жетілдірудің барлық кезеңінде оның бағдарламаларын жақсартып отыру қажет:

– корпусқа енгізілетін электрондық мәтіндерге филологиялық сараптама жасау;

– латын және төте жазу графикалық жүйесінде жазылған мәтіндерді кирил графикасындағы қазақ жазуына көшіру;

– омонимдерді автоматты түрде ажырататын бағдарлама жасау;

– метабелгіленім бағдарламасын әзірлеу. Бұл бағдарлама: корпус құрылысын жасауға; корпусты ақпаратпен толықтыру үдерісін қадағалауға; корпустың ауқымдылығы мен үйлесімділігін бағалауға; ақпаратты дәл іздеу мен сұрыптауды қамтамасыз етуге; сұралған ақпараттардың барлық контекстерін тауып беруге; сұралған ақпараттың жиілігі туралы толыққанды мәлімет беруге және т.б. мүкіндік бере алатындай жақсартылып отыруы тиіс. Мәселен, қазіргі ҚТҰК-да 23 параметр бойынша метабелгіленім бағдарламасы әзірленді: оның **9-ы мәтін параметріне** (1 - атауы, 2 - енгізілу мерзімі, 3 - сөз санына қатысты көлемі, 4 - мәтіннің қолданылу аясы туралы параметр – өндірістік-техникалық, оқыту-ғылыми, ресми-іскери, публицистика, жарнама, діни, тұрмыстық, көркем әдебиет және т.б.), 5 - мәтін тақырыбына қатысты параметр, 6 - хронотоп (оқиға орны мен мерзімі), 7 - мәтін типіне қатысты параметр: (публицистикада: шағын мақала, сұхбат, шолу, комментарийлер және т.б.; ресми-іскери тілдегі: заң, қаулы, кодекс, өмірбаян, акт, келісімшарт және т.б.), 8 - мәтін жанрына қатысты параметр: (тарихи проза, фантастика, оқиға, юмор, сатира, балалар әдебиеті, өмірбаяндық шығармалар және т.б.), 9 - мәтін стиліне қатысты параметр: **2-уі мәтін авторы** туралы параметр, **1-уі пайдаланушы аудиторияға** қатысты параметр, **5-уі қосымша** параметр. Алайда корпустық сайтта олардың барлығы іске қосылмай тұр. Сондықтан болашақта бұларды толық іске қосып, метабелгіленім параметрлерін әр ішкорпус спецификасына, функциясына сәйкес кеңейту маңызды.

– мәтіндерге автоматты түрде грамматикалық (морфологиялық) белгіленім қою бағдарламасын жетілдіру. Мысалы, ҚТҰК-да морфологиялық (әрбір тілдік бірлікке морфологиялық ерекшеліктеріне қарай белгі қою, нақты айтқанда, сөздердің сөз табына қатысы, сөзформалардың грамматикалық сипаттамасы (септік, көптік, шақ, рай және т.б. категориялық мағыналары); лексикалық (сөздің лексикалық мағыналарын көрсету), фонологиялық (сөздің құрамындағы тіл дыбысына/фонемаға артикуляциялық сипаттама беру), фонетикалық

(сөздің лингвалды сингармотембрін, яғни жуан/жіңішкелігін анықтау, буынға бөлу, әр буынның фонетикалық сипаттамасын беру) белгіленім енгізілгенімен, лексика-семантикалық белгіленім әзірлену үстінде де, әлі іске қосылған жоқ (тілдік бірліктің семантикалық сипаты бойынша белгі қою, мысалы, «адам», «зат», «кеңістік», «жылдамдық», «қозғалыс», «иелік» және т.б.).

– мәдени-семантикалық (лингвомәдениеттанымдық және этнографиялық) белгіленімдер енгізу бағдарламасы әлі де жетілдіруді қажет етеді;

– ҚТҰК стилистикалық белгіленім енгізу бағдарламасы қолға алынған жоқ;

– фонетикалық-просодикалық белгіленім енгізу бағдарламасы сөздер құрамындағы дыбыстық модификацияларды, яғни сегментті бірліктер өзгерісінің негізгілерін, ал суперсегментті өзгерістердің ішінде әуеннің синтагмалар тоғысындағы қозғалысын, кідірістерді ғана көрсетеді. Болашақта фонема инвариантының вариант, вариацияларының мейлінше толық қоры мен корпус бірлігіне екпін сияқты басқа да просодикалық сипаттарына қарай белгі қоюды жетілдіру қажет;

– синтаксистік белгіленім енгізу бағдарламасын әзірлеу де болашақ еншісіндегі маңызды міндет;

– омонимдер ажырататын бағдарламаның тілтанымдық негіздерін әзірлеу алға тартылып отыр;

– корпустық грамматикалық сөздіктер әзірлеу мен қазақ тілінің морфологиялық стандартын, іздеу механизміне негізделген метабелгіленімдер әзірлемесін жетілдіру жұмыстарын жалғастыру;

– қазақ тілінің негізгі стильдерін қамтитын теңгерімделген оқыту субкорпустары базасы негізінде жиілік сөздік әзірлеу; қазақ тілінің тек белсенді сөздерінен құрастырылған сөз тіркестері мен сөйлемдердің анықтамалығын (анықтамалық құралын) жасау;

– метабелгіленімдерді, морфологиялық, синтаксистік, лексика-семантикалық және мәдени-семантикалық белгіленімдерді толықтай автоматтандыру, лексикалық және грамматикалық омонимдерді автоматты түрде ажыратушы бағдарламасы, сондай-ақ автоматты іздеу тиімді механизмдерінің бағдарламасын жасау;

– интернет-ресурстарынан қазақ мәтіндерін автоматты түрде көшіріп алу парсинг бағдарламасын; қазақ сөздері мен сөз тіркестерінің жиілік сөздігін алу үшін сөз бен оның қоршауындағы тіркесім жиілігін құрастыру бағдарламасын жетілдіру;

– ҚТҰК-ның онлайн-режимде жылдам жұмыс істеуін қолдау мен оны жаңартып отырудың жүйелі жұмыстары; пайдаланушы мен корпустың байланысын орнататын корпустық менеджерлердің арнаулы бағдарламасын әзірлеу;

– ҚТҰК жасау үшін ғылыми лабораторияларды бағдарламалық-техникалық және технологиялық қамтамасыз ету: серверлік жабдық, жүйелік бағдарламалар, корпустық әзірлемелер бойынша дайындалған бағдарламалар орнату және т.б.; жасалған ҚТҰК-ны онлайн-режимдегі жұмысын тұтынушы талаптарына сәйкес үзбей жақсартып отыру қажет.

1.2 Қазақ тілінің ұлттық корпусын әзірлеудің тарихи кезеңдері

Ғылым мен білім берудің бүгінгі даму сатысында ақпараттық технологияларды қоғамның барлық саласында кеңінен қолдану ісі кезек күттірмейтін өзекті мәселелердің біріне жатады. Осы орайда Мемлекет басшысы Қасымжомарт Кемелұлы Тоқаев 2020 жылғы 1 қыркүйектегі «Жаңа жағдайдағы Қазақстан: іс-қимыл кезеңі» атты Қазақстан халқына Жолдауында: «Цифрландыруды ұлттық бәсекеге қабілеттілікке қол жеткізудің басты құралы» дей келе, «азаматтардың сауаттылық деңгейін, олардың цифрлық білімін арттыру қажеттілігін, сонымен қатар 2021 жылғы 1 қыркүйектегі «Халық бірлігі және жүйелі реформалар – ел өркендеуінің берік негізі» атты Қазақстан халқына Жолдауында: «Қазіргі заманда бәсекеге қабілеттіліктің басты факторының бірі – барынша цифрландыру. ... Біз еліміздің орасан зор ақпараттық-телекоммуникациялық әлеуетін іске асыра білуге тиіспіз», – деп атап көрсетті.

Мемлекет басшысының тапсырмаларын орындау мақсатында, Қазақстан Республикасының мемлекеттік тілі – қазақ тілінің мемлекеттік статусын, қоғамдық қызметін күшейту, қолданыс аясын кеңейту үшін оны әлемдік тіл білімінің ақпараттық технология кеңістігіне шығару қажеттігі күн тәртібінен маңызды орын алды.

XXI ғасырдағы ақпараттық технологияның қарқынды дамуы ғылымның барлық салаларын цифрландыруды қажет етті. Қазіргі кездегі әлеуметтік қызметтердің барлық түрі, хат алмасу, ақпарат алмасу және т.б. ұялы телефондар арқылы орындалатын заманда, қағаз формадағы кітапты оқудан гөрі, қажетті ақпаратты әлеуметтік желілерден іздеу үрдіс алды.

Осы ретте жаратылыстану салаларында ғана емес, гуманитарлық салалар өнімдерін де замана сұранысына сай әзірлеу әрі бүгінгі ғылым-білімнің жетістіктерінің арқасында адамдардың өмір сүру формасын жеңілдету, оңтайландыру мүмкіндіктерін қарастыру қажеттігі туды. Ұлттық корпустар жасау мәселесі де әлемдік тіл білімінде әсіресе осындай лингвистикалық зерттеулер жүргізу мен практикалық құралдар жасаудың таптырмас құралы ретінде XX ғасырдың аяқ шенінде күрт дамыған болатын. Осылайша, Корпус жасаудың аса маңыздылығы мен қажеттігі туындап, алғашқы болып, **1963 ж.** АҚШ-та **Браун корпусы** жасалды.

Бастапқыда бұл корпустың көлемі 1 млн. сөзқолданыстан тұрып, оның құрамында әрбіреуі 2 мың сөзқолданысқа тең 500 мәтін қамтылған. Алайда тілдік зерттеулерді сапалы жүргізу тек ауқымды тілдік материалдар негізінде ғана жүзеге асатындығы біліне бастады. Осыдан корпустарға қойылатын көлемділік шарты туындады.

Корпустық лингвистиканың ағылшын тіл білімінде кең өркен жаюын ғалымдар АҚШ-та компьютерлік техника мен XX ғасырдың 60-80 жылдары британ лингвистикасындағы интеллектілік ахуалдың белсенді дамуымен түсіндіреді. Осы кездерде тілдік зерттеулердің ең көп бөлігі компьютерленген мәтіндік корпустарға лингвистикалық талдау жүргізуге бағытталған-ды. Мұндай зерттеулердің нәтижесі 2001 жылы Бирменгем университетінде корпустық лингвистика кафедрасын ашуға және **International Journal of Corpus**

Linguistics (Корпустық лингвистиканың халықаралық журналын) баспадан шығарып тұруға мүмкіндік туғызды. Бірнеше тілдің материалдары бойынша құрастырылған, түрі мен қызметі жағынан ерекшеленетін корпустар және солардың негізінде неше түрлі **сөздіктер түзіліп, грамматикалар** жазылды [5], [6], [7].

Қазақ тіл білімінде Тәуелсіздік алуымызбен байланысты ғылым салалары да кең қанат жая бастады. Тіл білімі саласында тілді адам факторы тұрғысынан зерттеу тіл мүмкіндігін жаңа көкжиектерге көтерді. Антропоэлекттік парадигмадағы когнитивтік лингвистика, психолінгвистика, функционалды лингвистика салаларымен қатар тілді математика, информатика салаларымен сабақтастықта зерттейтін Қолданбалы/Компьютерлік лингвистика саласы да дами бастады.

Қазақ тіл білімінде статолінгвистика саласы 1970 жылдары белгілі математик ғалым Қалдыбай Бектаевтың жетекшілігімен және профессор Асқар Жұбанов, Аманкеш Зекенова, Алмасбек Белботаев және т.б. статист ғалымдарымыздың жиілік сөздіктер құрастыру тәжірибесінен басталады.

Институтымыздың Қолданбалы лингвистика бөлімінде көп жылдар бойы менгерушілік қызмет атқараған профессор Асқар Құдайбергенұлы Жұбанов Қолданбалы лингвистика саласының Статолінгвистика, Формалды модельдеу, Компьютерлік лингвистика сияқты салалары бойынша зерттеулер жүргізе жүріп, Қазақ тілінің ұлттық корпусын жасау қажеттігін ертеден-ақ көре білді.

Мұны ғалым А.Жұбановтың: «Әлемдік корпустық лингвистиканың даму сипаты – ұлттық толық мәтіндерді арнайы зерттеу нысаны етіп алуды қажет етуде. Қазақ мәтіндерінің стилдік, құрылымдық, мағыналық, функционалды және т.б. сипаттарын анықтау өзекті мәселелердің бірі болып отыр. Сондықтан автоматтанған қазақ тіліндегі мәтіндер корпусының компьютерлік базасын құру да ғылыми және практикалық жағынан аса құнды мәселе», – деуінен көреміз [8].

А.Жұбанов Қолданбалы лингвистика бөлімінің зерттеу бағытын осы мәселеге бұрды. Қазақ тілінің ұлттық корпусын жасау ісін орталықтандыру керек деген үндеу тастап, бұл туралы «Ана тілі» газетіне көлемді мақала жариялады және бұл туралы Ғылым комитетіне арнайы хат жолдаған болатын [9].

ҚТҰК жасаудың алғашқы қадамдары аз көлемдегі мәтіндерді базаға салып, морфологиялық белгіленімдерді мәтінге қолдап қою арқылы **С.Құлмановтың** жетекшілігімен жүзеге асырылды. Атап айтқанда, *2009-2011 жылдары*:

– қазақтың белгілі жазушылары А.Құнанбаев, М.Әуезов, Ә.Кекілбаев, М.Мағауин, М.Мақатаев шығармаларынан алынған таңдама мәтіндер (көркем стиль) бойынша мәтіндік бірліктерге (*150 000* сөзқолдананыс) **«қолдап» (ручная разметка)** морфологиялық белгіленімдер қойылып, корпус базасына енгізілді.

Осы аз көлемдегі база негізінде корпуста енгізілген басқа да мәтіндерге морфологиялық белгіленім жасалды [9]. Алайда «қолдап» енгізілген морфологиялық белгіленім қойылған база көлемі аз болғандықтан, мәтіндерге морфологиялық белгіленімдерді автоматты түрде қоя алатын морфологиялық анализатор жасау қажеттігі туды. Бұл морфологиялық анализатор корпуста енгізілген мәтіндерді автоматты түрде түбір мен қосымшаға бөліп, олардың

сөз табына қатысын және сөз формаларының түрленімін (грамматикалық/морфологиялық) сипаттау қажет еді.

Осы ретте «Қазақ әдеби тілінің аннотацияланған ұлттық корпусы» (2012-2014) атты тақырып ұсынылып, осы кезеңде ф.ғ.д., проф. А.К.Жұбанов, ф.ғ.д. А.Ә.Жаңабекованың ғылыми жетекшілігімен морфологиялық анализатор жұмысын жүзеге асыру үшін:

– 15-томдық «Қазақ әдеби тілінің сөздігінің» (2011 ж.) барлық томдары бойынша реестр сөздер біртөмдік «Қазақ сөздігіндегі» (2013 ж.) сөздермен толықтырылып және олардың сөз табына қатысын көрсетететін грамматикалық белгіленімдер тексеріліп, реттеліп, **қазақ тіліндегі негізсөздер сөздігінің (реестр) электрондық базасы** жасалды;

– қазақ тіліне тән әрбір сөз табы үшін құрастырылған сөзөзгертуші (сөз түрлендіруші) қосымшалардың компьютерлік базасы жасалды.

Мұнымен қоса базада сөзөзгертуші/түрлендіруші қосымшалар құрылымы алдын ала грамматикалық деңгейде белгіленді. Әрбір сөзөзгертуші/түрлендіруші қосымшаға қысқартылған түрде белгіленімнің шартты белгісі жазылды, яғни сөзөзгертуші қосымшаларға қысқарту арқылы грамматикалық белгіленімдері көрсетілді. Жекеленген сөз таптарының сөзөзгертуші/түрлендіруші аффикстерінің тізімі А.Жұбановтың негіз сөздер үшін соңғы дыбыстарының аяқталуына (дауысты, дауыссыз, қатаң, ұяң және т.б.) қарай құрастырылған кестелері бойынша жасалды;

– аталған негіз сөздер базасы мен түрленім қосымшаларының морфологиялық белгіленім базасы негізінде **туындатылған сөзформалар базасы** жасалды;

– негізсөздер (реестр) мен түрленім қосымшалар кестесі арқылы туындатылған сөзформалар базасы негізінде корпус мәтіндеріне немесе кез келген көлемдегі таңдама мәтінге автоматты түрде грамматикалық белгіленім қоятын **морфологиялық анализатор** жасалды [10].

– Ономастика бөлімі қызметкерлері шығарған қазақ тіліндегі антропонимдер мен топонимдер сөздіктері корпусқа енгізіліп, программа осы тізім бойынша корпус мәтіндерінде кездесетін онимдерге **антропоним** не **топоним** екендігін көрсете алатын болды.

Корпус, ең алдымен, лингвистикалық зерттеулер үшін қажетті көлемді массивті мәтіндер болып табылады. Зерттеу жұмыстарының белгілі бір жүйе бойынша жүргізілетіні белгілі. Өйткені тіл – қарым-қатынас құралы, қолданыс аясы өте кең күрделі жүйе болғандықтан, тілдік зерттеулер белгілі автор шығармалары бойынша немесе белгілі бір кезеңде жарық көрген көркем не баспасөз немесе тарихи ескерткіштер бойынша, сол сияқты белгілі бір жанр, стиль немесе белгілі бір тақырып, мәселе бойынша тарам-тарам болып жіктеліп кете береді. Осындай әртарапта шашылып жатқан материалдарды корпус жадына жай ғана енгізбей, белгілі бір жүйеге бағындырып енгізсе, зерттеу жұмыстарының да тиімді орындалуына орасан зор ықпал етеді. Осы ретте Қолданбалы лингвистика бөлімі корпусқа енгізілетін мәтіндерді метамәтіндік (экстралингвистикалық) ақпаратпен қамтамасыз ету қажеттігін ескере келе, әлемдік корпустардағы метамәтіндік ақпараттар енгізу тәжірибесін зерттеу

және оны қазақ тілінің корпусына енгізу мақсатында «Қазақ тілінің ұлттық корпусындағы метамәтіндік белгіленімдердің ұстанымдары» (2015-2017) атты ғылыми жобаны ұсынып, **6 миллион** сөзқолданыстан тұратын мәтіндерге метабелгіленімдер енгізді. Мәтіндер поэзия және қазақ тілінің 5 стилінен жинақталды.

а) **көркем проза** (1 миллион сөзқолданыс);

ә) **поэзия** (1 миллион сөзқолданыс);

б) **ғылыми-гуманитарлық мәтіндер** (1 миллион сөзқолданыс);

в) **публицистикалық мәтіндер** (1 миллион сөзқолданыс);

г) **ресми (іскери) стиль мәтіндері** (1 миллион сөзқолданыс);

д) **сөйлеу стилі** (1 миллион сөзқолданыс) т.б.;

– әлемдік тіл біліміндегі метабелгіленімдер қоюдың теориялық және практикалық әдіс-тәсілдері зерттеліп, оларды Қазақ тілінің ұлттық корпустарына енгізудің тиімді әдістері айқындалды; нәтижесінде **метабелгіленім қоюдың жартылай автоматты компьютерлік бағдарламасы (9 параметрмен)** жасалды [11];

– жоғарыда аталған 5 стильден таңдалып алынған мәтіндерге метабелгіленімдер (мәтін авторы; мәтін атауы; мәтіннің жазылған уақыты мен жері; мәтіннің көлемі; мәтін типі; мәтін дереккөзі т.б.) енгізілді;

– Қолданбалы лингвистика бөлімінің 6 миллион сөзқолданыстан тұратын қазақ тілінің 5 стилінен алынған «**Қазақ тілінің мәтіндер корпусы**» сайты (**89.250.84.132**) жасалды;

– зерттеу тәжірибесі бойынша «**Корпустық лингвистика**» (автор А.Жұбанов, А.Жаңабекова) **монография (ISBN 978-601-7293-43-7) /оқу құралы** жарық көрді [12].

– фонетикалық белгіленімнің буынға бөлу нұсқаулығы жасалып, ізделген **сөзді автоматты түрде буынға бөлетін компьютерлік бағдарлама** корпусқа енгізілді [13].

2015 жылы Тіл білімі институтының бір топ қызметкерлерінің қатысуымен «Қазақ тілінің ұлттық корпусын қалыптастырудың негізгі факторлары, Қазақстан Республикасындағы тілдердің үштұғырлы идеологиясына сәйкес коммуникативтік және тұтастырушылық қызметін арттыру» атты ғылыми жобасы бойынша корпус жұмыстарын жетілдіру бойынша төмендегідей нәтижелер алынды:

– онбестомдық «Қазақ әдеби тілінің сөздігіндегі» **реестр сөздер түсіндірмелері/дефинициялары корпус базасына енгізілді**; Осы түсіндірме сөздік базасы арқылы Корпуста ізделген сөздің лексикалық мағыналары беріліп тұрады.

– Қолданбалы лингвистика бөлімінде орындалған морфологиялық белгіленімдер «қолдап» енгізілген мәтіндер базасы негізінде **морфо-семантикалық белгіленімдер (дара, күрделі, түбір, туынды, салт, сабақты, т.б.) қою бағдарламасы** жасалды;

– корпустан ізделген сөзді түбір мен қосымшаға ажырататын (лемматизация) және автоматты түрде морфологиялық және морфо-семантикалық белгіленім қоятын, сонымен қатар сөздер дефинициясы берілетін «**Қазақ тілінің**

қанатқақты корпусы» атты сайт (А.Фазылжанованың мұрындық болуымен) жасалып, іске қосылды.

2016 жылы А.М.Фазылжанова Э.Сүлейменованың кеңесшілігімен «Қазақ тілінің ұлттық корпусын» жасаудың **концепциясын** жазып, министрлікке жолдады [14]. Нәтижесінде:

2016 жылы Тіл білімі институтының сол кездегі директорының ғылым жөніндегі орынбасары ф.ғ.к. **А.М.Фазылжанның** ғылыми жетекшілігімен «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» ғылыми жобасы Тіл білімі институты және әл-Фараби атындағы Қазақ ұлттық университетінің «Жалпы тіл білімі және еуропа тілдері» кафедрасының оқытушыларының (ғылыми кеңесші проф. **Э.Д.Сүлейменова**) қатысуымен жүзеге асырылды. Бұл кезеңде:

– Қазақстандағы электрондық мәтіндер базасы бар мекеме, ұйымдардан **үлкен көлемде мәтіндер жинақталды;**

– стильдер бойынша **метамәтіндік белгіленім қоюдың лингвистикалық нұсқаулығы мен 23 параметрден тұратын компьютерлік бағдарламасы** жасалып, корпусқа енгізілді;

– лексика-семантикалық белгіленім бағдарламасын жасауға қажетті **қазақ тіліндегі сөздердің лексика-семантикалық топтарының тізімі жинақталды** (профессор **Б.Момынованың** жетекшілігімен), қазіргі кезде осы классификация арқылы Институтта етістік сөз табының семантикалық топтары жіктеліп, лексика-семантикалық белгіленімдер корпусқа енгізуге дайындалып жатыр.

«*Латынграфикалы жаңа ұлттық әліпби негізінде қазақ жазуын жаңғыртудың ғылыми-лингвистикалық базасын әзірлеу*» атты бағдарламалық-нысаналы зерттеу аясындағы 3.2 тапсырма (тапсырмаға жауапты: А.Ә.Жанабекова) (Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі қазақ тілінің корпусық ресурстарын жаңа латын графикасында әзірлеу және жасау) бойынша:

2018 жылы – 2 млн. сөзқолданыстан тұратын көркем мәтіндер;

2019 жылы – 2 млн. сөзқолданыстан тұратын оқулық мәтіндері;

2020 жылы – 2 млн. сөзқолданыстан тұратын ғылыми мәтіндер және 2 млн. сөзқолданыстан тұратын публицистикалық мәтіндер жоспарланып, Қолданбалы лингвистика бөлімінің ғылыми қызметкерлері корпус базасына 2018 жылы 23 параметрмен метабелгіленім қойылған 2 млн. сөзқолданыстан тұратын көркем мәтіндерді, 2019 жылы әртүрлі ғылым салалары бойынша пән оқулықтары мәтінін, 2020 жылы 2 млн сөзқолданыстан тұратын публицистикалық мәтіндер мен 2 млн. сөзқолданыстан тұратын ғылыми мәтіндерді жинап, метамәтіндік белгіленімдер енгізді.

Бұл кезеңде корпус қазақ жазуын латын әліпбиіне көшіру арқылы реформалау бойынша бекітілген акут негізді латын әліпбиімен конвертор бағдарламасын қолданып, кирилше мәтіндерді автоматты түрде латынға көшіріп, латын қарпімен де іздей алды. Келешекте Корпусты латын графикасына да ауыстыруға мүмкіндіктер жасалды.

ҚТҰК жасау ісі Институттағы ғылыми жобалар аясынан бөлек, магистрлік диссертациялар деңгейінде де жүзеге асырылды.

Магистрлік жұмыстар аясында (Диссертация жетекшісі: А.Ә.Жаңабекова):
1) сөзжасамдық белгіленім (Пірманова Күнсұлу) әзірлемесінің сынама нұсқасы жасалды [15].

Корпуста сөзжасамдық қосымшалар базасы арқылы анализатор туынды түбір сөздердің жасалу тәсілдері мен жасалу жолдары туралы ақпарат береді. Алайда кейбір жағдайда омоним, омограф формалардың, яғни сөзтүрлендіруші қосымшалармен, сол сияқты сөзтудырушы омографтармен астас келіп, қате көрсету мәселесі әлі шешіліп жатыр.

2) Магистрлік диссертация (А.Мурсал) аясында қазақ тіліндегі **күрделі етістіктер сөздігінің** (түсіндірмесі берілген) базасы жасалған [16]. Алдағы уақытта бұл база арқылы морфологиялық анализатор күрделі етістіктерді жеке сыңарларға бөлмей, біртұтас бірлік ретінде сипаттайтын болады.

Фразеологизмдер белгіленіміне қажетті **фразеологизмдер базасы** І.Кеңесбаевтың «Фразеологиялық сөздігі» мен «Қазақ әдеби тілінің түсіндірме сөздігі» материалдары негізінде жасалды. Фразеологизмдер/тұрақты тіркестер базасы арқылы конкордансқа шыққан сөйлемдердің ішінен фразеологизмдер айырықшаланып және тінтуірмен көрсеткенде, мағынасы шығып тұрады.

А.Мейірбекова (ғылыми жетекші – А.Фазылжан) корпустағы омонимдерді ажырату мәселесі бойынша магистрлік диссертация қорғады [17]. Осы жұмыста анықталған **омоним ажырату алгоритмдері** алдағы уақытта корпус бағдарламасына енгізуде бастам болмақ.

2020 жылы Тіл білімі институтының директоры **А.М.Фазылжанның** басшылығымен корпус сайтының дизайны мен интерфейсі толық жаңартылып, **qaz.corpus** домен атауымен қалың көпшілік пайдалануына ұсынылды. Бүгінде сайт ауқымды корпустық ресурс ретінде көпшілікке қолжетімді тегін форматта үздіксіз жұмыс істеп тұр.

16.08.2020 – Әрбір ішкорпуста метабелгіленімдер бойынша іздеу жүйесі жетілдірілді.

19.08.2020 – Компьютерлік бағдарламаның түбір сөздерді ғана емес, олардың түрленген формалары бойынша да мысал табу мүмкіндігі қарастырылып, сөзформадан іздеу ұяшығы қосылды.

25.08.2020 – Ізделген сөз бойынша табылған деректер статистикасын шығару мүмкіндігі қамтамасыз етілді.

26.11.2020 – Пайдаланушыға қажетті сөзді іздеу бойынша мәтінді көшірудің программалық жолдары жасалды.

2020 жылы грамматикалық қосымшалар бойынша іздеу бағдарламасы жасалды.

2021-2022 жылдарға арналған «Мемлекеттік тілді латынграфикалы ұлттық әліпбиге көшіруді қамтамасыз ететін жаңартылған нормативтік сөздіктер мен академиялық басылымдар сериясын әзірлеу» атты бағдарламалық нысаналы қаржыландыру жобасы бойынша да ҚТҰК-ның мәтіндік базасы Қолданбалы лингвистика бөлімінің міндеттері негізінде толықтырылды.

2021 жылы публицистика мен көркем мәтіндер бойынша бөлім қызметкерлері мен институттағы өзге қосымша орындаушылар көмегімен **7 миллион-**

нан астам сөзқолданыстан тұратын мәтіндерге метабелгіленім енгізіліп, корпус базасына салынды.

05.10.2021 – Корпус мәтіндерінің сөзформаларының реестр тізімі және реестр сөздер бойынша іздеу (поиск) жасалды.

2022 жылы Институт «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» (2022-2023 жж.) атты бағдарламалық нысаналы қаржыландыру жобасын (ғылыми жетекшісі – **А.Фазылжан**) орындауды қолға алды. Бұл жоба ҚТҰК-ның ішкорпустарын жасауға мүмкіндік берді. Бұл ретте:

18.10. 2022. – «Қазақ тілінің аймақтық сөздігі» материалдары негізінде Қолданбалы лингвистика бөлімінің орындауымен **Диалектілік ішкорпус** әзірленіп, корпусқа салынды (сөзқолданыс саны – 350 мың), (ішкорпус жетекшісі – **А.Жаңабекова**);

05.05.2022 – Терминология және Қолданбалы лингвистика бөлімінің қызметкерлері қатысуымен көркем әдебиет және ісқағаз стильдері бойынша қазақша-орысша мәтіндер теңестіріліп, әзірленген **Параллель ішкорпус** ҚТҰК базасына енгізілді (сөзқолданыс саны – 2 миллион), (ішкорпус жетекшісі **С.Құлманов**);

10.10.2022 – Тіл тарихы бөлімі қызметкерлері араб, латын графикасымен жазылған XI–XX ғғ. жазба мұралары бойынша **Тарихи ішкорпус** әзірледі. (сөзқолданыс саны – 1 миллион). (ішкорпус жетекшісі – **А.Сейтбекова**);

02.05.2023 – **Қазақ тілінің ауызша корпусы** Ахметтану, Тіл мәдениеті және Фонетика бөлімі қызметкерлерінің қатысуымен әзірленді. Бұл ретте қазақ тілінің ауызша тілінен аудиожазбалар жинақталып, олардың сөзжазбасы қағазға түсірілді (орфография), сондай-ақ сөйлеуші және норма орфоэпиясы беріліп, оларға академик **З.Базарбаеваның** кеңесшілігімен **просодикалық белгіленімдер** енгізілді. (сөзқолданыс саны – 1 миллион). (ішкорпус жетекшісі – **Қ.Күдерінова**);

Ауызша сөз мәтіндеріне орфографиядан орфоэпияға автоматты түрде аударым жасайтын «**Орфоэпиялық конвертор**» бағдарламасы іске қосылды;

18.05.2023 – Оним түрлері бойынша іздеу жасайтын және лингвомәдени ақпараттар берілген **Ономастикалық корпус** әзірлемесі жасалып, Ұлттық корпус базасына енгізілді (ішкорпус жетекшісі – **Қ.Рысберген**).

2023-2024 жылдарға арналған «**Мәдени-репрезентативті және жарнама мәтіндерінің ішкорпустарын әзірлеу**» атты бағдарламалық нысаналы қаржыландыру жобасы бойынша (ғылыми жетекшісі – **А.Жаңабекова**):

28.05.2023 – **Қазақ тілінің мәдени-репрезентативті мәтіндер ішкорпусы** жасалып, онлайн қызметі іске қосылды;

29.05.2023 – **А.Фазылжанның** идеясы бойынша Этнолингвистика бөлімі әзірлеген **мәдени-семантикалық белгіленімдер** Лингвомәдени корпус базасына енгізілді. Бұл ретте **Н.Уәлидің** мәдени семантиканы ашу әдістемесі басшылыққа алынды. Мультимедиялық корпус бастамы болатын бейнежазбалар (видеоролик) салынды. (ішкорпус жетекшісі – **Э.Өтебаева**);

20.05.2023 – Академик **Ә.Қайдардың** этнолингвистикалық түсіндірмесі берілген мақал-мәтелдер мен осы әдіспен Институт магистранттарының

«Ұлттық паремиологиялық қор: генезисі және прагматикасы» пәнін оқу барысында әзірлеген **Мақал-мәтелдер ішкорпусы** ҚТҰК базасына қосылды. (ішкорпус жетекшісі – **А.Жаңабекова**);

13.09.2023 – **Ахмет Байтұрсынұлы мәтіндерінің ішкорпусы** әзірленіп, Ұлттық корпус базасына енгізілді. (Ахметтану бөлімі, ішкорпус жетекшісі – **Қ.Күдерінова**);

25.05. 2023 – Қазақ тіліндегі көше, көлік, сауда орындары, Бақ, интернет жарнамалары мәтіндері қамтылған **Жарнама корпусы** әзірленді. (ішкорпус жетекшісі – **Г.Исаева**).

2023-2024 жж. «**Қазақ мәтінін автоматты тану: лингвистикалық модульдер мен IT-шешімдер әзірлемесі**» атты бағдарламалық нысаналы қаржыландыру (ғылыми жетекшісі – **А.Фазылжан**) жобасы бойынша:

28.09.2023. – **Оқу корпусының** алғашқы сынама нұсқасы әзірленді (ішкорпус жетекшісі – **А.Жаңабекова**).

Қазіргі кездері ҚТҰК дизайнын, метамәтіндік және лингвистикалық белгіленімдерін жетілдіру жұмысы үзбей жүргізіліп отыр. Корпус – ашық жүйе болғандықтан, оның мәтіндік базасын толықтыру, белгіленімдерін түрлендіру және жетілдіру, іздеу жүйесін күшейту сияқты кезекті жұмыстар тоқтамауы тиіс.

Қазақ тілінің жоғарыда аталған кейінгі жылдары жасалған ішкорпустары туралы толық ақпаратты осы монографияның алдағы тарауларынан оқитын боласыздар.

2 АУЫЗША ІШКОРПУС

2.1 Ауызша ішкорпус: сөйлеу тілінің ерекшеліктері туралы

Сөйлеу тілі – әдеби тілді тұтынушылардың дайындықсыз, машық бойынша бейресми жағдайда, тыңдаушымен бетпе-бет сөйлесетін сөзі, әдеби тілдің функционалдық тармақтарының бірі. Сөйлеу тілінің нормалары грамматикада, сөздіктерде тіркелмейді, сондықтан сөйлеу тілі нормалары әдеби тілдің кодификацияланбаған нормасына жатады. Әдеби тілдің нормалары кодификацияланған стильдік тармақтарына сөйлеу тілі осы ерекшеліктерімен қарама-қарсы қойылады. Бейресми қарым-қатынаста (сөйлеу тілінде) жазбаша, яғни кітаби тіл элементтерін қолдану сөз мәдениеті тұрғысынан дұрыс деп танылмайды. Сондықтан сөйлеу тіліне тән ерекшеліктерді жазба тілден айыра білу – сөздің коммуникативтік сапасына қойылар басты талаптың бірі.

Сөйлеу тілінде *боп, кеп, бересашы, күн жауат, бартын, барат, айтат, кеп қапты* т.б. дыбыстық құрамы жағынан ықшамдалған бірліктер жиі қолданылады. Кодификацияланған мәтінде бұл тәрізділер қатеге есептелгенмен, сөйлеу тілі үшін норма. Сондай-ақ сөйлеу мәтінде әлгідей сөздердің толық нұсқасы да қолданылады. Алайда жалпы әдеби нормаға тән бірліктердің қолданылуы сөйлеу тілінің статусына нұқсан тигізбейді.

Сөйлеу тілінің тағы бір тілдік ерекшеліктерінің бірі – тұрпатмеже тұрғысынан ықшамдалған бірліктердің мәтінде ұзына бойы қолданылуы шарт емес, сондықтан ықшам нұсқа мен толық нұсқа алма-кезек қолданыла береді. Бұл қасиет сөйлеу тіліне тән еркіндікті білдіреді: *кеп кетті/келіп кетті, жел соғат/жел соғады*.

Сөйлеу тіліне тән нышанды (маркерлі) бірліктермен қатар жалпы әдеби тілдік нормаға тән бірліктер де қолданыла береді. Сөйлеу тілінде кітабилік нышаны бар элементтерді қолдану сөйлеу тілінің нормасына жат деп есептеледі. Мысалы, *-уда (оқушы жастар білім алуда), шылық-шілік, егін шаруашылық, мал шаруашылық, бірінен саналады, болып табылады* т.б. кітаби тілдік нышаны бар элементтер.

Сөйлеу тілінде *ұстаз, қаламгер* тәрізді көтеріңкі мазмұндағы бағалауыш сөздердің қолдану мүмкіндігі әлдеқайда төмен. Кітаби нышаны бар мұндай сөздер сөйлеу тілінің табиғатымен үйлеспейді, сөзді жасандылау көрсетеді.

Ақиқаттағы заттарды функциясына қарай жалпылама атау – сөйлеу тілі мәтінде жиі кездеседі: *Кесетін қайда кеткен?* («қайшы» деген мағынада); *жазатын бірдеме бар ма?* («қалам» деген мағынада); *отыратын ештеңе таппадым* («орындық» деген мағынада).

Ақиқаттағы заттарды түр-түсіне қарай жалпылама атау сөйлеу тілі мәтінде жиі қолданылады: *қызылыңды кигенің дұрыс* («көйлегіңді» деген мағынада). Денотатты белгілі бір бөлігімен атау сөйлеу тілі лексикасында жиі кездеседі. *Бір үйге бір доңғалақ керек екен* («автомашина» деген мағынада); *аспиринге барып келейін* («дәріханаға барып, аспирин алып келейін») т.б.

Болымсыз мағынаны білдіретін синтаксистік құрылымды болымды мән-де айту сөйлеу тілі мәтінде жиі кездеседі: *Шай ащы екен* («ішпеймін» деген мағынада).

Сөйлеу тілінде сөйлемнің ықшам, қысқа беймәңделген, редуцияланған түрлерінің кездесуі контекст, сөз жағдаяты (ситуациясы), аялық білім тәрізді құрылымдармен тығыз байланысты. Мысалы:

А: – *Қашан келдің?*

Б: – *Кеше.*

Сен, мен деген мүшелердің беймәңделуі контекст арқылы айқындалып тұр. А. мен Б-ның кездесуі – сөз жағдаяты (ситуация). Б-ның Астанаға кеткенін А. біледі. Бұл – аялық білім.

– *Сағат бір болды, кете беріңдер!* Контекст – сөз субъектісі екінші жақ; Сөз жағдаяты – «компьютер басып отырған қыздар»; Аялық білім – түскі үзіліс (сағат 13), «асхана».

Көпқұрамды күрделі сөздер сөйлеу мәтінінде бір ғана сөзбен айтылуы жиі кездеседі: – *Жиналыс болды ма? («ата-аналар жиналысы болды ма?»).* *Сәбізді қойдың ба? (сәбіз шырынын ішіп жүрсің бе?).* Сонымен бейресми жағдайда қарым-қатынастың ресми түріндегідей сөйлеу, солай сөйлеу, яғни «кітаптағыдай» сөйлеу қатеге саналады. Сондай-ақ қалай сөйлесе, солай жазу кодификацияланған мәтінде қате деп танылады.

Сөз субъектісі, яғни сөйлеуші өзін тұлға ретінде көрсетеді. Оның сөйлеген сөзінен белгілі бір этносқа, ұлтқа тән екені, мәдениет деңгейін, дүниетанымдық ерекшелігін, этикалық және құндылықтар жөніндегі түсінігін байқатып тұрады.

Сөйлеу тілі диалог формасында жүзеге асады. Диалогтің ойдағыдай болуы, ойдағыдай аяқталуы үшін қарым-қатынас жасауды қажетсіну – негізгі шарттардың бірі. Бұл қажетсіну коммуникативтік мүдделілік деп аталады.

Коммуникативтік мүдделілік деңгейіндегі қатынаста диалогке қатысушылар арасында олардың әлеуметтік мәртебесіне, әлеуметтік рөліне қарамастан паритеттілік орнайды.

Мұндай қарым-қатынас жасауға деген мүдделілік, тілдесуге деген құлшыныс таныстыққа, етене жақындыққа, әлеуметтік тәуелділікке, эмоционалдық фонға (тілектестік, бейтараптық, ұнатпаушылық) қарамастан, диалогтің тепе-тең дәрежеге өтуіне әсер етпейді.

Құлшыныс танытушылардың сөйлесу, тілдесу және келісімге келуі тілдік қарым-қатынастың нәтижелігін көрсетеді.

Әңгімелесушінің жан дүниесіне байланысты түсіністік, сөйлесушілердің, тілдесушілердің дүниетанымындағы жақындық, өмір тәжірибесіндегі ұқсастық, мүдделестік, мәдени, туыстық т.б. апперцепциялық ая – диалогтің нәтижелі болуына әсерін тигізетін шарттардың бірі.

Ишара, ым арқылы, интонация, реплика, тон, дауыс тембрі арқылы өзара түсіністікке жол ашылып, тілдік қарым-қатынастың ойдағыдай болуы үшін тыңдаушы сөз мәніндегі интенцияны аңғаруы, сөйлеушінің не айтқаны емес, нені білдіргісі келетінін аңдауы керек. Сөйлеушінің коммуникативтік интенциясын тыңдаушының аңғара білуі диалогтің ойдағыдай өтуін қамтамасыз етеді. Коммуникативтік интонация вербалды деңгейге дейін, яғни сөз тыңдаушыға жеткенге дейін қалыптасады.

Әдетте тыңдаушы айтылып жатқан сөз ағымынан қалыспай, айтылған сөздің мәнін ой елегінен өткізіп үлгереді де, сөйлеушінің коммуникативтік

тұспалын (интенциясын) «шешеді». Бұл жерде тілдік, психологиялық, биологиялық, физиологиялық процестер асқан шапшаңдықпен қатарлас жүреді. Сөйлеу (речевая деятельность) теориясы бойынша әдетте а) логикалық құрылым мен тілдік құрылым толық сәйкес келмейді; ойдың сөзбен айтылмайтын құрылымдары бар; ә) мән-мағынаны айқын аңғартатын тәсілдер бар да, мән-мағынаны айқын көрсетпейтін тәсілдер бар, тікелей аңғартатын тәсілдер бар да, тұспалмен аңғартатын тәсілдер бар.

Мән-мағына үзіктерін айқын көрсетпейтін тұстар да мән-мағынаны тұтас күйінде емес, ішінара білдіретін тұстар да сөйлеу тілі үшін қалыпты жағдай. Бұл тыңдаушы мен сөйлеушінің арасындағы күрделі қарым-қатынасты көрсетеді. Сөз тұспалын түсінуде ойға «қысым» түсетін кез осы тұста. Сөйлеу тілі осынысымен қызық, осынысымен ұтымды. Олай болмаған жағдайда сөйлеу тілі тым сұрқай, бірөңкей, тыңдаушыны мезі етіп, жалықтырар нәрсеге айналар еді. *Шөлдеп кеттім, пысынап кеттім* десе, күннің ыстықтығын білдіреді немесе *күн ыстық екен* дегені *шөлдеп кеттім* дегенді білдіреді. *Күн тымырсық екен* деген сөз *жел соқпады* дегенді, я болмаса саяжайға бару керектігін аңғартады.

Тілдік қарым-қатынастың мазмұны ондағы барлық тілдік элементтердің мағынасынан әлдеқайда кең. Сондықтан тек тілдік бірліктердің мағынасын біліп қою жеткіліксіз және олардың мағынасын білумен шектелу қарым-қатынастың ойдағыдай болуына кепіл бола алмайды. Өйткені сөздің мазмұндық құрылымында вербалды түрде берілетін мағынадан басқа бейвербалды мағыналық компоненттер болады.

Әрине, мағыналық блоктар белгілі бір тілдік бірліктер арқылы, олардың тіркесі арқылы санада қалыптасады. Тілдік бірліктердің комбинациясы сөйлеушінің аялық білімін, пікірін, танымын, ой өрісін, эмоциялық күйін, тілекталабын т.б. білдіреді. Бұл – сөйлеушінің тілдік компетенциясына, құзыретіне байланысты.

Тілдік құзырет ұғымы – коммуникативтік өзара әсер етудің орталық ұғымы. Тілдік құзырет тілді білу дегенге саяды, ол тілді білу дегеніміз тиісті ойды әртүрлі тілдік амал-тәсілдермен еркін жеткізе білу; белгілі бір тілде айтылған сөздің мәнін біліп, сентенциясын, тұспалын аңғару; ұқсас тұлғалардың мән-мағынасын ажырата білу; сырт тұлғасы әртүрлі құрылымдардың, бірліктердің ортақ мазмұнын аңғара білу; құрылымы дұрыс сөйлемдерді тілдік тұрғыдан кәте сөйлемдерден ажырата білу.

Коммуникативтік құзырет (компетенция) тілдік қарым-қатынастағы әлеуметтік мәдени нормаларды және стереотиптерді білу дегенге саяды. Коммуникативтік құзыреті жоғары адам әртүрлі деңгейдегі бірліктердің мағынасы мен сол элементтердің өзара тіркесіндегі мағынасын білумен бірге мәтін деңгейіндегі әлеуметтік өлшемдерді, мысалы, тіл қатудың әртүрлі формаларын қолдана алады. Белгілі бір оқиға өзінің бағасын шынайы түрде бере отырып, оған тыңдаушы тарапынан жауап реакция туғыза алады. Қарым-қатынасты интимдендірудің тілдік әдіс-тәсілдерін қолданудың ретін біледі. Бұл тұста сөйлеушінің әртүрлі коммуникативтік ситуацияда паронимдер, фразеологизмдерді, мақал-мәтелдерді, мәтіндік клишелерді, прецедентті мәтіндерді

т.б. білуінің мәні айрықша. Бұлар – тілдің әлеуметтік нормаларын меңгерудің жоғарғы көрсеткіштері.

Сонымен, тілдік қарым-қатынас ойдағыдай болу үшін сөйлеуші тыңдаушыға тек жалаң фактілерді ғана айтып қоймау керек, тек ақиқатты айту жеткіліксіз, сонымен бірге жанама ақпараттар беру қажет. Сөз ішіндегі, сөз арасындағы ирония, юмор, символ, образ, парадокс т.б. тыңдаушыны бейжай қалдырмайды, тілдесушілердің өзара түсіністігіне жол ашатын жанама тәсілдерге жатады.

Бетпе-бет айтылған сөздің мәні, дауыс кідірісі, сөздің қарқыны, сөйлеушінің бет-әлпетіндегі құбылыстар, ым мен ишара бетпе-бет сөйлесуде сөз мазмұнын, құрылымын толықтырып отыратын маңызды элементтер болып табылады.

Қарым-қатынастың тілдік дүниеліктердің қалыптасуына көп жағдайда тілдік емес факторлар ұйтқы болады. Тілдік емес факторлар көзқарас, әрекет, жай-күй, эмоция, білім, сенім т.б. мәнділіктерден тұрады. Осымен байланысты тілдік қарым-қатынастың ойдағыдай болуы немесе сәтсіз аяқталуы сөйлеушінің тек тілдік формаларды талдауына тәуелді емес. Диалогтің ойдағыдай өтуі лингвистикалық және лингвистикалық емес факторларға байланысты.

Диалогтегі сәттілік пен сәтсіздік а) коммуниканттардың әлеуметтік-стереотиптік ерекшеліктеріне; ә) коммуниканттардың аялық біліміне, б) коммуниканттардың құзыретіндегі жақындыққа, я болмаса алшақтыққа; в) коммуниканттардың жыныс, жас, тұлғалық ерекшеліктеріне байланысты жайттарды ескеру/ескермеуден болуы ықтимал.

Диалогтің өрбуіне немесе үзілуіне коммуниканттардың көңіл күйі де ерекше әсер етеді; бөгде адамның қатысуы да диалогтің өрбуіне кері әсерін тигізуі мүмкін. Бөтен коммуникативтік орта сөйлеушілер үшін қолайсыздық туғызады. Тілдік қатынас өз деңгейінде болмайды да, сөйлеушілердің «шығармашылық» қабілеті шектеліп қалады.

Қарым-қатынастың тепе-теңдігі (паритеттілігі) сақталмаған жағдайда диалог ойдағыдай өрбімей қалады. Сөйлеушілердің біреуі басым болады да тақырып тізгінін қолдан шығармайды, әріптеске сөз бермейді. Айтып отырған сөзіне тыңдаушының қандай қабақ танытып отырғаны, дұрыс, не бұрыс түсініп отырғаны т.б. жағдаяттарға мән бермейді, сөйтіп, сөйлеуші коммуникацияға қатысушының көңіл күйін аңғаруға қабілетсіздігін көрсетеді.

Диалог сөздің өрбуіне, нәтижелі болуына кесірін тигізетін нәрсе орынсыз ескерту жасау, сөзді орынсыз бұзу, орынсыз сұрау қою, орынсыз көп сөйлеу, айтылғанды орынсыз қайталау т.б. Мұндай «орынсыздықтар» тыңдаушымен есептесудің, қарым-қатынастағы ынтымақтастықты (кооперативность) бұзып, сөздің берекесін кетіреді.

Сөйлеушілердің мәдени-әлеуметтік ерекшеліктерімен байланысты айтылмыш сөз орынсыз болып, коммуникативтік сәтсіздікке әкеліп соқтыруы мүмкін.

Сөйлеушінің тілді білу деңгейі тым төмен болған жағдайда тілдік қатынас сәтсіздікке ұшырауы мүмкін. Мұндайда сөйлеуші жергілікті сөздерді қолдануы мүмкін. Қайсыбір сөйлеушінің тыңдаушы біле бермейтін арнаулы терминдерді

орынсыз қолдануы, сондай-ақ шетелдік сөздерді орынсыз жұмсауы диалогтің сәтті өтуіне кедергі жасайды.

Сөйлеу тілінің әдеби түрі, сөйлеу тілінің қарапайым түрлері. Олардың нормалары туралы.

Коммуникативтік мақсаттың типтеріне қарай, әріптестердің қатысу амал-тәсілдеріне қарай, олардың атқаратын рөлдері арасындағы ерекшеліктеріне, репликалық сипатына қарай тілдесу жанрлары әңгімелесу (беседа), сөйлесу (разговор), әңгіме айту (рассказ); ұсыныс, өтініш, талас, ескерту, ақыл-кеңес, хат-күнделік, жазба (записка) күнделік болып бөлінеді.

Әңгіме. Тілдесудің бұл түрі кооперативтік стратегияда белгілі бір мәселе бойынша пікір алысу дегенге саяды. Әріптестер өздерін қызықтыратын жайттар жөнінде мәлімет алмасады; белгілі бір мақсат көздемей-ақ, пікір алмасады, әртүрлі жаңалықтарды сөз етеді. Белгілі бір мәселе бойынша пікір алысу кезінде сөйлеушілер өзінің көзқарасын білдіруде әлеуметтік-мәдени стереотиптерді басшылыққа алады.

Тілдесу жанрының екінші бір типінде сөйлеушілер көбіне бір-біріне қошаметсөз (комплимент) айтады, біріне-бірі сенім білдіреді. Бірін-бірі шынайы танитындығын көрсетеді. Әңгіменің мұндай типіне аксиологиялық модальдылық тән.

Әңгіме жанрының үшінші бір түріне мейрам, жиын-той, мереке жағдайындағы қарым-қатынас жатады. Өзіл-қалжың, күлдіргі сөздер айтылады. Сөйлеушілер қызықты жайттардан сөз қозғайды. Әңгіменің мұндай типіне эмоционалдық модальдылық тән. Модальдық мағына көбіне, *қызық екен, байқайсыз ба, солай ма, шынымен солай екен-ау, кім білсін, қайдан білейін* т.б. реттеуіштер арқылы беріледі.

Сөйлесу. Сөйлесуде ақпарат алмасады; өтініш білдіреді, бұйрық береді, талап етеді, ақыл-кеңес береді, иландырады, ұсыныс айтады.

Тілдесу. Екеуара қатынастағы ұрыс-керіс, дау-жанжалды, қақтығыс, кінәлауды т.б. айқындауды мақсат етеді. Сөз арасында *екеуіміз түсінісіп алатын жайт бар, сенімен әңгіме бар* т.б. сөз орамдары айтылады. Лидер көбіне сұрау қояды. Реплика-сұрау түрінде болады. Сөйлеушілердің мәдени-әлеуметтік нормаларды, этикеттік нормаларды ұстана білуі диалогтің нәтижелі болуына септігін тигізеді.

Талас. Талас немесе сөзталастыру – шешім қабылдау я болмаса ақиқатты айқындау мақсатындағы пікір алысу. Белгілі бір мәселеге байланысты көзқарас әртүрлі болады да, сөйлеушілер өзара тілдесуге мүдделілік танытады. Бірақ сөзталастың бұл фазасы тілдік формалар арқылы объективтенбейді. Этикеттік нормалар, мәдени-әлеуметтік нормалар сақталады. Аргумент шынайы болуға тиіс.

Пікірталастың мақсаты ақиқатқа жетіп, бір шешімге тоқтау болып табылады.

Пікірталас барысында эпистемалық модальдылықтың қалыптасуы мүмкін. Тілдесушілер арасында көзқарастың, ұстанымының ортақ жақтары атап айтылады. Өзінің пікірін аспанға көтеретін декларация коммуникативтік қатынастың сәтсіздігіне саяды.

Белгілі бір мәселеде келіспейтіндігі жөнінде сөйлеушілер өзінің аргументтерін, дәлелдерін ұсынады. *Бәлкім, былай ойлаған дұрыс болар, мәселені басқаша қырдан да қарауға болатын сияқты. Сіздің пікіріңізді сыйлаймын, бірақ мен бұл жерде басқаша ойлаймын* т.б. тәрізді сөз орамдары, реттеуіш сөздер қолданылу тиімді деп есептеледі. Қалыптасқан қағида бойынша кісінің жеке басының сапасын, артықшылығын сөз етпейді.

Белгілі нәрсе, айдан анық, арасы жер мен көктей, әрине тәрізді реттеуіш сөздерді қолдану тәжірибеде жиі кездеседі.

Сөзталас кезінде психологиялық қысым жасау, білмейтінді бетке басу пікірталастағы диалогтің сәтсіз тұстары болса, сөйлеуші жақтардың бір-біріне құрметпен қарамауы, мәдени-этикалық нормалардың бұзылуы болып табылады.

Әңгіме (расказ) – сөйлеу тілі жанрларының бірі, монологтік сипаттағы сөз. Тілдесудің басты стратегиясы – кооперация, келісім, ынтымақтастық. Сөйлеушінің біріне өзінің ақпарат түріндегі коммуникативтік діттемін жүзеге асыруға мүмкіндік беріледі. Тақырып ақауы сөйлеуші куә болған уақиға, факт төңірегінде болады. Немесе өзге адамның куә болған уақиға, фактісі әңгімеленуі ықтимал.

Әңгіменің жекелеген фрагменттері бір-бірімен байланысты болып, ақпарат біршама тұтас, бірізді баяндалады. Ара-тұра тыңдаушының (тыңдаушылардың) мақұлдаған, танданған, қызықтаған репликаларына мүмкіндік беріледі. Айтушының адресаттың аялық біліміне негізделген эпистемиялық межесі болады.

Әңгіменің тақырыбы мен сипаты күлкілі, қорқынышты, тағылымдық мәні бар уақиға, факті болуы сөздің модальдығын көрсетеді. Әңгімедегі фразеология, идиоматика, прецедентті мәтіндер бір жағынан, мағыналық блоктар болса, екінші жағынан сөйлеушінің сөз шеберлігін танытатын құрылымдар болып табылады.

Тарих. Әңгіменің тарих түрінде сөйлеуші іс-әрекеттің иесі, іс-қимылдың орындаушысы ретінде көрінбейді. Тарихтың коммуникативтік мақсаты бұрынғы өткен жайттан мағлұмат беріп қоймайды, мағыналық қорытынды шығарады.

Тарих әңгімеден сөйлеу синтаксисінің барлық ерекшеліктерін кездестіруге болады. Тематикалық үзіктер, ассоциативті шегіністер, эллипсистік құрылымдар, лексикалық элементтердің экспрессиялығы, баяндаудың дайындықсыз өтуі, нақтылауыш лексемалардың көптігі, қыстырма сөздердің жиі ұшырасатындығы, баяндау барысында сөйлеушінің тыңдаушыны назарға алып отыруы тарихи әңгіменің стилистикасын көрсетеді.

Хат. Жақын адамға арналып жазылады. Адресанттың ой-сезімі жазба формаға түседі. Хат авторы өзін тұлға ретінде табиғи түрде экспликациялаудың мүмкіндігін пайдаланады. Бұл – барша тілдік қарым-қатынастың ең басты прагматикалық шарты. Хатта сәлемдесу, қоштасу формулалары сақталады.

Сөйлеу тілі жанрларына жазбасөз (записка), күнделік жатады.

Тілдік қарым-қатынас этикеті. Сөз әдебі. Тілдесудің ойдағыдай өтуінің этикалық шарты бар. Адресантқа жалпы оң пейіл білдіру, тілдесуге ықыласты

екенінді көрсету, оның сөзін түсіністікпен ден қоя тыңдау т.б. сондай-ақ әңгімеге, сөзге құлақ түріп, назар аударудың вербалды белгілері реттеуіш репликалар арқылы жүзеге асады.

Сонымен бірге ым, ишара, жылы жүз таныту т.б. бейвербалды таңбалар арқылы да сөзге өзінің көңіл қойып, назар аударғанынды көрсетуге болады.

Сөз этикеті – моральдық нормалар, мәдени-ұлттық дәстүрлер негізінде қалыптасқан сөз әдебінің ережелері.

Этикеттік нормалар арнайы этикеттік сөз формулалары арқылы жүзеге асады. Ең басты этикалық ұстаным – сөздің, әңгіменің ұзына бойында – «аман-дасудан» бастап «қоштасуға» дейін тепе-теңдікті паритеттілікті сақтау.

Этикеттік қаратпа атауыштар сөздің, әңгіменің барысында ерекше өң береді. Сөйлеушілер әлеуметтік рөліне қарай, бір-біріне алыс-жақындығына қарай «сіз» немесе «сен» деп сөйлейді. Этикеттік қаратпа атауыштар оларда байланыс орнату, сөйлеушілердің бірін-бірі жақын тартуына дәнекерлік қызмет атқарады. Әңгіме барысында этикеттік қаратпа атауыштарды бірнеше мәрте қолданып отыруға тура келеді. Мұның өзі сөйлеушіге, я болмаса тыңдаушыға деген жылы ықыласты, бір-бірінің сөзіне ден қойғандықты танытады.

Ата-ананың, әсіресе ананың балаға деген мейірімі, еміренісі *шырағым, қарағым, қоңыр қозым, құлыным* т.б. тәрізді этикеттік қаратпа атауыштардың эмоциялық мағынасы ерекше әсерлі болып тұрады. Ал баланың ата-анаға қаратып айтар этикеттік қаратпа атауыштар, ағайын-туыстар арасында айтылатын этикеттік қаратпа атауыштар өз алдына бір-бір жүйе.

Қазақ қоғамындағы мәдени-ұлттық дәстүрде бөгде адамдарға айтылатын этикеттік қаратпа атауыштардың да ерекше жүйесі бар: *мырза, ана, апай, қарындас, тәте, аға, інішек, бикеш, айым, төре* т.б. Бұл қаратпа атауыштар бөгде адамның жас, жыныс т.б. әлеуметтік ерекшеліктеріне байланысты қолданылады.

Ертеректегі қазақ қоғамында адамдардың қоғамдағы орнына, әлеуметтік мәртебесіне байланысты қолданылатын этикеттік қаратпа атауыштардың қалыпты жүйесі болған: *Хан, ханым, ханшайым, ханша, ханзада; бек, бегайым, бекбике, бекзада, бике; би, биеке, бише; бай, байеке, бәйбіше; патша, патшайым, патшазада; төре, сұлтан, тархан, әмір, мырза* т.б. әлеуметтік мәртебесі жоғары, лауазымды адамдарға және оның жақын туыстарына қатысты айтылған. Осылардың ішінде ақсүйектерге қатысты айтылатын *хан, ханзада, ханым, ханшайым, бек, сұлтан, төре* тәрізді лауазымды мәртебені білдіретін сөздерді кісінің есіміне қосып айту қатаң этикеттік норма болған. Оны сақтамау адамға ат шапан айып кесіліп заң жолымен жазаланатын болған.

Бірақ мәдени-этикеттік нормаларға байланысты кейбір құндылықтар кестік дәуірде таптық идеологиясы тұрғысынан қаралып, сөзқолданыстан ығыстырылды, соның салдарынан мәдени-ұлттық нормалардың кейбір түрлерін қолдануда жалғастық, сабақтастық болмай қалды. Әсіресе көпшілікке қарата айтылатын этикеттік қаратпалар жүйесінде *жолдастар* атауышын қолдану совет қоғамында мәдени этикалық нормалардың біріне айналды.

Қазіргі кезде сөйлеуші көпшілікке қарата айтылатын сөзде *мырзалар мен ханымдар* деген этикеттік қаратпа атауыштарды қолданады. Әрине, бұрынғы

қазақ қоғамындағы жиындарда, көпшілік орындарда мәртебелі лауазымды орынды қолдануға ерекше көңіл бөлген: ханды *хан*, әйелін *ханым* деп атаған. Жалпы әйел атауын *ханым* деп айтпаған. Шаршытопқа *уа халайық, әлеумет* деп, лауазымдыларға *игі жақсылар* деген этикеттік қаратпа атауышты қолданған.

Қазіргі қоғамда жағдай басқа. Шаршытопта айтылатын сөзде әйел қауымын, ерлер қауымын ерекше қаратпа атауыштармен атауға тура келеді. Осымен байланысты *ханымдар мен мырзалар* деп сөйлеу мәдени этикеттік дағдыға айналып келеді. Бұл жерде *мырза* қаратпа атауышы орынды көрінеді. *Мырза* – тек әлеуметтік артықшылыққа ие топтың өкілін ғана емес, жалпы ер адамды құрметтеу нышанымен айтылатын этикеттік мән алған атауыш.

Ал шаршы топта әйел жынысты адресаттарға *ханымдар* деген атауышты қолдану, біздің байқауымызша, пәлендей орынды емес, ұлттық-мәдени нормадан ауытқуға жатады. Әсіресе ресми жиындарда, дипломатиялық кездесулерде *айымдар мен мырзалар* деп қолданудың жөні басқа тәрізді. Қазақ қоғамында әйел жыныстыға *айым* деп, ерекше құрметпен қараған. *Айым* сөзі әйел жыныстыға қарата айтылатын этикеттік қаратпа атауыш қызметінде жұмсалған. *Патша, хан, бек* тәрізді әлеуметтік мәртебесі жоғары лауазым иелерінің әйел туыстары *патшайым (патша-айым), ханшайым (ханша+айым), бегайым (бек+айым)* түрінде, *айым* сөзімен қосарлана айтылған. Осыған қарағанда құрметтеу, дәріптеу мәні ерекше атауышты *айым* атауышына шаршытопта айтылатын этикеттік қаратпа атауыш етіп қолдану *ханымдар* қаратпасына қарағанда әлдеқайда лайық көрінеді.

Этикеттік формулалар. Қазақ тілінде аса жиі кездесетін, әлеуметтік мәдени мәнділігі жоғары, коммуникативті ниет танытатын тілдік формулалар мен әдіс-тәсілдер қалыптасқан. Солардың біралуаны амандасу, қоштасу, кешірім сұрау, өтіну, ризалық білдіру, құттықтау, тілек білдіру, көңіл айту, бата беру т.б. Бұлардың ішінде бата беру, көңіл айтудың рәсімдік сипаты бар.

Этикеттік формулалар – коммуникативтік құзыреттіліктің аса маңызды бөлігі. Этикеттік формулаларды жақсы білу, коммуникативтік ситуацияға лайық етіп, орнымен жұмсай білу – сөйлеушінің сөз мәдениеті, жалпы мәдениетінің жоғары екенін көрсететін, сонымен бірге қандай әлеуметтік топтың мүшесі екенін байқатып тұратын белгі.

Сөздің эвфемизмделуіне сөйлеуші ерекше мән береді. Тілдік қарым-қатынас кезінде мәдени ахуалды бұзбау, әріптестің көңіліне келетін, жанына жағымсыз тиетін сөзді айтпау эвфемистік номинацияны қолдануға, сөздің тігісін жатқызып (эвфемизмде) қолдануды қажет етеді.

Қоғамда қалыптасқан мәдени-ұлттық стереотиптерді бұзатын, жағымсыз ассоциация туғызатын, тыңдаушыға ауыр тиетін т.б. сөздерді, сондай-ақ тыйым салынған сөздерді перифрастикалық номинациямен атаудың қазақ тілінде қалыптасқан тілдік әдіс-тәсілдер әр алуан.

Бұл тәрізді перифрастикалық атаулар көбіне адамның табиғи кемшіліктеріне, қайтыс болуына, жыныс мүшелеріне, физиологиялық күйіне т.б. байланысты болып келеді. Бұлардың ішінде рәсімделген (ритуалданған) эвфемистік номинация ерекше орын алады.

Сондай-ақ бір нәрседен бас тарту, өтінішін орындамау тәрізді жайттарға тура жауап түрінде емес, жауап сөз жұмсартылған түрде беріледі: *уақытым болмай тұрғаны, мүмкін болмай тұр; қуана-қуана барар едім* т.б.

Этикеттік норма бойынша адамның көзінше *мынау, ол* деп үшінші жақпен айту сыпайылыққа жатпайды.

Байырғы этикалық-мәдени норма бойынша ауыл тұрғындарының арасында жасы үлкен әйел адамдарды *Әміржанның көкесі-ау, Құтжанның әкесі-ау* т.б. деп, күйеуін үшінші жақта айтылатын қаратпамен атаған. Бұлар – эмоционалдық реңкі жоғары этикеттік таңбалар.

Айтылып жатқан әңгіме аса эмоционалды болып, сөз арасында сөйлеушінің сөзін кейде бөлуге, сөз қыстыруға тура келеді. Тыңдаушы белсенділік танытып, сөйлеушінің сөзін қоштайтынын, растайтынын, мақұлдайтынын фактілермен айтуға, көрсетуге тырысады. Сөйтіп, өзінің сөйлеушіні толығымен қостайтынын білдіретін, *сөзіңіз аузыңызда, кешіріңіз, сөзіңіздің жаны бар, сөзіңізді тірілтіп жіберейін* т.б. тәрізді «жұмсартулар» арқылы сөзді бөлу жағдайы кездесіп отырады. Мұндай тәсілдер коопертивті стратегияға қайшы емес. Ал *тұра тұрыңыз, оның жайы олай емес, бұлай, менің білуімше* т.б. деп сөзге тосыннан жасалған килігулер арқылы сөзді бөлу бейкоопертивті стратегияға жататын тәсілдер. Бұлайша мәдени-этикеттің бұзылуы сөзге килігушінің коммуникативтік құзыретінің төмендігін көрсетеді, жалпы сөз мәдениеті деңгейінің төмендігін байқатады.

«Сен» деп сөйлеу мен «сіз» деп сөйлеу. Қазақ қоғамында қалыптасқан мәдени ұлттық стереотип бойынша, *сіз* бен *сеннің* әлеуметтік этикеттік мәнділігі ерекше. *Сіз* – жасы үлкен адам мен бейтаныс адамға ресми жағдайда айтылады. Адамдардың әлеуметтік рөлі қатаң ескеріледі. Ал «сіз»-бен сөйлеудің қалыптасқан нормасын бұзып, тыңдаушыға «сен»-мен сөйлеу оның әлеуметтік мәртебесін кеміту болып табылады.

Жасы қарайлас, әлеуметтік мәртебесі шамалас бөтен адамдар таныса келе «сен» деп сөйлеуге көшеді. Қазақ мәдени дәстүрінде «сен» мен «сіз» қатаң нормаға жатады. *«Сен деген – сөздің анығы, сіз деген – сөздің сынығы»* деген мәтел этикеттік мәні бар бірлікті орынды жұмсау керектігіне меңзейді. Қазақ мәдени дәстүрінде «сіз» деп сөйлеу коопертивтік стратегияда тек сөздің сынығын эвфемизмделуін ғана білдірмейді. Аталмыш сөздердің басқа да этикалық-мәдени мәні бар. *Сіз* сөзі коммуникативтік ситуацияға қарай ресмилік мағынаны білдіріп, кітаби реңк алады.

Кейде тәжірибеде, «сіз»-дің осы тұстағы кітабилік реңкі ескерілмей жататыны байқалып қалады. Мысалы, мектеп оқушыларына «сен» деп сөйлеу керек пе? әлде «сіз» деп сөйлеу керек пе? деген мәселе. Мектеп оқулықтары, байқап отырсақ, «сен» деп сөйлеу үрдісіне көшкен. Бұл жайтты мақұлдас бұрын, я терістемес бұрын, алдымен оқулықтың әлеуметтік тіл мәртебесін еске алайық. Оқулық тілінің нормасы әдеби тілдің кодификацияланған нормасына, ресми қатынаста қолданылатын кітаби тілдік құрылымға жатады. Мұнда оқушыны тілмен баурау, жақын тарту тәрізді коммуникативтік діттем, әрине, сөз стратегиясына байланысты интимдендірудің тілдік элементтері де ұшырасуы мүмкін. Бірақ бұлай ету сөйлеу тіліндегідей тым еркін емес. Сондықтан оқулық тілінің

ресмилік, кітабилік сипатын естен шығармау керек, бұл – бір. Екінші, әсіресе, бастауыш сыныптарда баланың дағдыланған тілі сөйлеу тілі болғандықтан, алғашында бала оқулық тілін мүлде жатсынбасын деген ниетпен сөйлеу тілінің жазбаша имитацияланған формалары да оқулық тілінде болуы ықтимал, алайда бұл айтылғандар барлық сыныптар оқулықтары *сен* деп сөйлеуге тиіс деген сөз емес. Жоғары сыныпқа барғанда оқушы *сіз* сөзінің сыпайылық қана емес, ресмилік мағынасын да ұғынуға тиіс. Жоғары сынып оқулықтары *тапсырманы орындаңыз, мазмұндама жазыңыз, әңгімелеп беріңіз* деп, *сіз*-бен сөйлеуге көшуі керек. Мәдени-ұлттық дәстүрде *сіз*-дің этикеттік орны ерекше. Оқушыларды коммуникативтік құзыреті деңгейінің жоғары болуына осы бастан көңіл бөлу қажет. Қазіргі кезде тілдік табиғи ортада «сен»-нің басым болып бара жатқанын көрсететін теріс құбылыс бар екенін байқауға болады. Оқушы сыпайылықты тек отбасынан үйреніп қоймайды, мұғалімнен, оқулықтан да үйренеді. Жоғары сынып оқушылары «сіз»-бен сөйлесудің ынтымақты стратегиялық мәні бар: мұғалім тілдік қарым-қатынаста тепе-теңдікті (паритет) сақтайды; оқушының өзіндік тұлға екенін сезінуіне мүмкіндік береді. Жоғары сынып оқушыларына арналған оқулықтар «*сіз*»-бен сөйлеуге, ал мұғалімдер «*сіз*»-ді коммуникативтік бағдарға алуға тиіс.

«*Сіз*» деп сөйлеуде қазақ мәдениетіне тән әлеуметтік ерекшелік, психологиялық дистанция, ресми қарым-қатынас тәрізді мәнділіктер бар.

Қорыта айтқанда, сөйлеуші өзінің сөзі арқылы өзіндік тұлға екенін таныта алуы, коммуникативтік мүдделілік, адресатқа икемделу, ал тыңдаушының сөйлеушіге бейімделіп, коммуникативтік діттемін түсінуі, сөз ситуациясына байланысты тілдік, тілдік емес қатаң талаптарды орындау – диалогтің ойдағыдай болуының шарты.

Тілдік қарым-қатынастың ойдағыдай болуы – сөйлеушінің тілдік құзыретіне және коммуникативтік құзыретіне тәуелді. Тіл құзыретіне грамматиканы білу, сөз мағынасын ажырату, белгілі бір мазмұнды әртүрлі тілдік тәсілдермен жеткізе білу, соның бірі перифраза т.б. жатады.

Коммуникативтік құзыретке осы аталғандардан басқа, мәдени-әлеуметтік нормаларды, сөз әдебіндегі ұлттық стереотиптерді білу, тілдік фактілерді коммуникативтік ситуацияға орнымен жұмсай білу т.б. жатады.

Сөйлеуші белгілі бір ақпарат беруге өзінің тіл дербестігін таныта білуі – тілдесудің ойдағыдай өтуіне әсерін тигізетін эстетикалық факторлардың бірі.

Тілдесуде мәдени-әлеуметтік стереотиптерді білу, аялық білімнің жеткілікті деңгейде болуы, тыңдаушының аялық білімін ескеру т.б. жайттар тілдік қарым-қатынастың ойдағыдай өтуіне әсерін тигізеді.

Сөз стратегиясы, сөз тактикасы, модальдылық сөйлеушілердің коммуникативтік бағдарына, діттеміне бағынышты.

Сөзге оң беретін, бедер беретін тілдік құралдарға анафора, антитеза гиперболола, литота, синонимдер тізбегі, градация, қайталау, эпитет, риторикалық сұрақ, метафора, метонимия, астарлы сөз, аллюзия, перифраза, субъективті модальдылық, қыстырма сөздер мен сөйлемдер т.б. жатады. Сөзге эстетикалық рең береді. Бейресми дискурста сөз этикасына, ережелеріне ерекше мән беріледі. Өйткені сөз этикасы – сөйлеушінің тыңдаушыға деген ілтипатын,

ықыласын танытатын, диалогтің ойдағыдай аяқталуына септігін тигізетін қажетті алғышарттардың бірі.

Сонымен, сөйлеу тілі дискурсында сөзді эллипсистендіру, жалқының орнына жалпыны, жалпының орнына жалқыны қолдану, дейксистердің жиірек жұмсалуды, экспрессивті сөздердің талғаммен қолданылуы (бұлар коммуникативтік тұлғаның вербалды-семантикалық сипаты) адресаттың әлеуметтік бейнесін тани білуі, кодқа салынған мәдени-ұлттық этикалық ақпараттарды ашуы (коммуникативтік тұлғаның когнитивтік сипаты), белгілі бір сөздерді қарым-қатынастың диттеміне сай жұмсалуды (коммуникативтік тұлғаның прагматикалық сипаты), коммуникацияға қатысушылардың тұлғасын айқындай түседі.

2.2 Қазақ тілінің ауызша корпусы және ауызша тіл

Қазақстанда қазақ тілінің ұлттық корпусы енді қолға алынып жатқан жылдары осы саланың мамандары «корпус құрастыру ісін орталықтандырып, үлкен күшпен мемлекет тарапынан қолдау көрсетілетін болса, мынадай корпус түрлерін жасап шығуға болар еді.

1) Қазақ тілінің қазіргі кездегі (немесе кезеңдік) бұқаралық ақпарат құралдары (газет, журнал бетіндегі) мәтіндерінің жеке корпусы;

2) Қазақша сөйлеу тілі жазбасының (мәтінінің) жеке корпусы (орыс тілінің «Корпус живой русской речи» тәріздес);

3) Қазақ тілінің мультимедиалық корпусы (корпустың негізін мәтіндердің бейне және дыбыс жазбалары құрайды);

4) Қазақ тілімен параллель тілдердің жеке корпусы (түркітілдес және үндіеуропа тілдері), мысалы, қазақ-қырғыз, қырғыз-қазақ, қазақ-өзбек, өзбек-қазақ және т.б., сол сияқты, қазақ-орыс, орыс-қазақ, қазақ-украин, украин-қазақ және т.б. қатар тілдер корпусы;

5) Қазақ тілінің поэтикалық мәтіндерінің жеке корпусы (орыс тілінің «Корпус русских поэтических текстов» тәріздес);

6) Қазақ тілінің білім беру корпусы (орыс тілінің «Обучающий корпус русского языка» тәріздес)» [12, 60 б.] деп, нақты санамалап, бағдарлама түрінде ұлттық корпус түрлерін жоспарлаған болатын.

Соның ішінде, қазақ тілі ұлттық корпусының маңызды бір құрамдас бөлігі ауызша сөздің корпусы болып табылатыны сөзсіз. Оның себебі ауызша тіл – жазба тілге қарағанда аса мобильді, жанды, қоғам мен органың өзгерістерін, сол тілде сөйлеушілердің санасы мен құндылықтарындағы дамуды, сөз мағыналарының өзгерістерін, үдерістерін көрсететін, жаңа лексемалар тудыратын еркін, тәуелсіз, бірақ тұтынушыға ең жақын коммуникация құралы және алаңы. Ауызша тілдің сапасы, ауызша тілдегі өзгерістер жалпы ұлттық тілдің дамуы мен жетілуіне әсер етеді. Әдеби тілдің нормаларын бұзады, коммуникацияны икемді ететін жаңа нормалардың жасалуына негіз болады.

Сондықтан тілдің ұлттық корпусы ауызша тілдің корпусынсыз толыққанды бола бермейді, тілдің шынайы, табиғи көрінісін бере алмайды.

Алайда тіл тілде ауызша тілдің корпусын жасаудың қиындығы бар. Сондықтан ауызша тіл корпусы барлық корпусстардың аз бөлігін қамтиды. Мы-

салы, «орыс тілінің Ұлттық корпусы құрамында 2008 жылдың қаңтарында бар болғаны 3,9% ауызша мәтін болды. Онда бұқаралық сөйлеу (публичная речь) (64,3%), бұқаралық емес сөйлеу (непубличная речь) (8,1%), кино сөзі (27,6%)» болды дейді А.Жұбанов, А.Жаңабекова [12, 57б.]. Британ ұлттық корпусында (BNC) ауызша тіл корпусы – 10% -ын ғана алады [18]. Ал түрік тілінің 1990 жылдан бермен қарайғы кезенді қамтитын ұлттық корпусында (TNC) ауызша корпус бар болғаны 2%-ын құрады.

Ал француз тілінің ұлттық корпусындағы ауызша мәтіннің синтаксистік және стилистикалық белгілеуіне қатысты артықшылықтары мен мүмкін болатын кемшіліктері бағаланатын зерттеулер жарық көріп жатыр [19].

Сондай-ақ британ корпусына енген ауызша мәтіндердің стихиялық құрылымын қабылдауды жеңілдету логикасын қарастыратын зерттеулер бар. Мақала авторының айтуы бойынша, 2012-2016 жылдары Ұлыбританияның түкпір-түкпірінен келген L1 Британдық ағылшын тілінде сөйлейтіндер арасындағы 11,5 миллион сөзден тұратын Британдық Ұлттық корпус пайда болған [20].

Ал қазақ тілі ұлттық корпусында әлі ауызша тілдің корпусы жасақталмаған.

Сол себепті А.Байтұрсынұлы атындағы Тіл білімі институтының 2022-2024 жылдарға арналған мақсатты қаржыландырылатын «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту Интернет-ресурсы» атты жобада алғаш рет ауызша тілдің ішкорпусын жасау жоспарланып отыр.

Осыған орай, тақырып орындаушылары – Тіл білімі институты қызметкерлері, алдымен,

- ауызша тіл дегеніміз не?
- ауызша тілдің мәтіндеріне нені жатқызамыз?
- ауызша тілдің мәтіндерін жинау тәсілі қандай?
- ауызша тілдің жазбаға түскендегі тілдік жүйе мен құрылымға қайшы келетін «жүйесіз», «нормасыз» сипатынан ішкорпусқа қажетті деректерді қалай шығарамыз?

- ауызша тілдің түрлері қандай деген сұрауларға жауап іздеуден бастады.

Сол себепті бұл мақаланы жазудың негізгі мақсаты қазіргі қазақ ауызша тілінің түрлерін, оның ішінде ауызекі сөйлеу тілінің ұлттық ерекшелігін сипаттау болды.

Зерттеу корпусқа салынуға тиісті ауызша тіл материалдарын игеру, ауызша сөйленген сөздердің транскрибациясын, стенограммасын, сөзжазбасын алу-дан басталды. Ауызша тіл деп айту және сөйлеу арқылы жарыққа шығатын, тыңдайтын аудиторияға бағытталған тілдік коммуникацияны алдық. Ауызша тіл, алдымен, әлеуметтік орта мүшелерінің коммуникативті діттеміне қызмет ететін басты тілдік құрал ретінде алынды.

Ауызша тіл - уақыт пен кеңістік жағынан мобильді, айтылған сөзді сол мезетте сол қалпында қайталау мүмкіндігі жоқ тіл [21]. Ауызша тілде, ең алдымен, сөз экономиясы болады [22]. Ол: коммуниканттардың паралингвистикалық, экстралингвистикалық, суперсегментті тәсілдер қорына: интонация, жест, тон, темп, мимика, затты көрсете алу, имплицитті семантика, фонға ие болуынан;

конситуация коммуникацияның толық мүшесі болуы, қоғамдық-әлеуметтік контекстің болуынан [23]; осыған байланысты сөздердің семантикалық байланысының күштілігі, морфологиялық формасыз, шылауларсыз байланыса беруінен [24]; анализизм, инверсияның күштілігінен; коммуниканттардың санасында вербалданбайтын жасырын мазмұнның болуынан және көбінесе аялық білім қоры ортақ коммуниканттардың сөзі болып табылатынынан шығады [25].

Ауызша тілде коммуниканттардың бір-бірін көріп отыруы, мобильділік, қайта реакция жасау мүмкіндігі оның құрылымына әсер етеді. Сондықтан бір сарынды баяндау болмайды, сөз тіркестері қарапайым, коммуникантқа әсер етуші функциясы динамикалық түрде ашық болады, автокоментатор, сөз қайталау, хезитация жиі пайдаланылады. Бір сөйлемде бірнеше рет қайталанған бір сөз интонациясы, тембрі әртүрлі болатындықтан, оны тыңдаушы әртүрлі семантикалық бірлік ретінде қабылдайды [26].

Ауызша тіл жазба тілден тез. Сондықтан ішкі сөйлеу кодына келген ой сол қалпында сыртқа шығады: ауызша тілдің синтаксистік құрылымы тілдік тәсілдерге қарағанда логикаға құрылады. Себебі ауызша тілде коммуниканттың өз еркілігі, өз меншігіндегі сөйлеу мәнері, коммуникативтік актіге дайындықсыз кірісіп кетуі бар [27]. Ауызша тілде жаңа сөздер, грамматикалық формалардың жаңа жұмсалымы көп болады.

Ауызша тілді тілтұтынушысының қолдануына қарай таныс ортада: 1) ауызекі сөйлеу тілі, таныс-бейтаныс ортада: 2) сөйлеу тілі немесе бейтарап сөйлеу тілі, 3) бейтаныс ортада: ауызша әдеби тіл (ауызша ғылыми тіл, ауызша іскери тіл, ауызша кәсіби тіл, діни сөз, ауызша шаршысөз, ауызша көркемсөз) деген түрлерін бөлеміз.

Ауызекі сөйлеу тілі – таныс ортадағы тұрмыстық қатынаста қолданылатын, тілдік қарым-қатынастың ең ілкі және өте өзгерімпаз түрі.

Сонымен, Ахмет Байтұрсынұлы атындағы Тіл білімі институтында жасалып жатқан ұлттық корпус құрамына институт директоры Анар Фазылжанованың бастамасымен бірнеше ішкорпустар енгізіліп жатыр. Олардың бірнешеуі мыналар: тарихи мәтіндер ішкорпусы, ономастикалық атаулар ішкорпусы, ауызша тілдің ішкорпусы, параллель мәтіндер ішкорпусы, мәдени бірліктер ішкорпусы және т. б.

Ауызша тілдің ішкорпусы қазіргі қазақ тілінде сөйленген ауызекі, сөйлеу стильдерінде, шаршысөз, көсемсөз, ғылыми сөз, іскери сөз стильдерінде жарыққа шыққан, аудиосы бар сөздердің қоры жинақталған ішкорпус болып табылады. «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту Интернет-ресурсы» атты мақсатты қаржыландырылған зерттеу аяқталғанша 1 млн сөзқолданыстан тұратын ауызша мәтіндердің жазбаға түскен орфографиясы, сөйлеушінің орфоэпиясы және нормадағы орфоэпия көрсетілген жазбалар және аудио нұсқасы салынған корпус жасалуы жоспарланды.

Шетел ғалымдары ауызша корпустың екі типі бар деген: бірі – дыбысталған мәтін (кітапты дауыстап оқу, оқылған жаңалықтар мәтіні), екіншісі – машықсыз сөз, яғни дайындықсыз сөйленген сөз (екі не одан да көп адамның әңгімесі, бір адамның ауызша сұхбаты) [28].

Жұмыс басталғанда алдымен осы типтердің қайсысы корпусқа алынады деген сауал тұрды. Өйткені алынатын сөздер белгілі бір мақсатқа сай болуы қажет. Ал қазақ ұлттық корпусында ауызша тілдің корпусы болуының мақсаты қазақ тілінің қолданыстағы күйін тану, зерттеу, меңгертуде есте болатын тұстардың әдістемесін шығару деп ойлаймыз. Өйткені ағылшын тілін үйретуде ауызша корпусты пайдалану тәжірибесі ертеден қолданылады [29]. Сонда ауызша тілдің ішкорпусына ауызша машықсыз сөздер, ауызша аудиторияға қызмет еткен барлық стильдегі сөздер алынды. Ал жазылған мәтін, оқылған, жазба тілмен сөйленген сөздер алынбады.

Осы тұрғыда ауызша корпустың метабелгіленімі жасалды.

Ең алдымен, сұхбат беруші, яғни сөйлеушінің аты-жөні берілетін болды. Яғни ішкорпусқа авторсыз емес, авторлы, тұлғасы бар сөздер алынды. Өйткені аты-жөні берілген сөздің басқа да параметрлері беріле алады. Мысалы, сұхбат берушінің жынысы, сұхбат берушінің туған жылы, мамандығы ауызша сөзді зерттеушілер үшін қажетті ақпараттар мен параметрлер болып табылады. Әсіресе сөздің лингвоәлеуметтік факторы сөздің сапасы мен ерекшелігіне тікелей әсер етеді. Сол себепті сөз авторының тұратын аймағы (қала тұрғыны, ауылдық жердің тұрғыны екендігі, Қазақстанның батысы, Қазақстанның шығысы, Қазақстанның оңтүстігі, Қазақстанның солтүстігінен екендігі немесе қазақ тілінің аймақтық түрі болып табылатын Баян-Өлгей аймағы, Шыңжаң өлкесі, Түркия не Ресейде тұрған болуы) міндетті түрде талдауға алынады. Метабелгіленімде көрсетіледі. Сонымен бірге сөз авторының жазба тілді білу, білмеуі, өзге тілдерді қолдану деңгейі, мысалы, таза қазақтілді болуы немесе екітілді, не көптілді болуы, сондай-ақ жас, жасы үлкен қария кісі болуы сөз сапасына әсер етеді.

Метабелгіленімде сөздің сөйленген уақыты, сөздің ұзақтығы берілуі – сөздің уақыт өлшемі бойынша өзгеру сипатын, норма мен дағдыға қатысты анықтаулар жасауға мүмкіндік береді. Мысалы, Байтұрсынұлы атындағы Тіл білімі институтының ауызша тіл ішкорпусында классик жазушылар Ғ.Мүсірепов, М.Әуезов, С.Мұқановтардың ауызша сөздері бар. Бұл сөздер М.Әуезов атындағы Әдебиет және өнер институтының қолжазбалар қорынан алынды. Осындай құнды дүниелерді сақтап келген институт басшылығына алғысымыз шексіз.

Сонымен, бірге сөйленген сөздің мынадай тақырыптары қамтылды: өмір-баян, тарих, білім, мәдениет, саясат, дін, әдебиет, дәстүр, тіл білімі, ғылым және т.б.

Ауызша сөздің конситуациясы, яғни сөз жағдаяты маңызды [30]. Сондықтан сөздің телеканал, телефон, уатсап желісі, инстаграмм, фейсбук, әлеуметтік желі каналы, жиналыс мінбері, дәрісхана, мерейтой, ас-садақа, отбасылық мереке (той, тілеу), шай үстінде, театрда, студия, конференцияларда сөйленген барлық түрлері алынды.

Және ең бастысы сөйленген сөздің стильдері оның қарапайым-тұрмыстық, сөйлеу, шаршысөз стилі, көркемсөз стилі, ғылыми стилі, іскери стиль, діни стиль болуы да зерттеушілер үшін жаңа тұжырымдар әкелетін нысандар болып табылады. Сондай-ақ оқылған мәтін, баяндама түрінде дыбысталған ауызша сөздер де ішкорпусқа алынады.

Ауызекі сөйлеу тілінің өзіндік мынадай ерекшеліктері байқалды:

а) сөз жағдаяты маңызды;

ә) жеке стиль болады;

б) нақ осы шақта түсінікті болады;

в) сөз синкретті сипат алады, жағдаят пен сөз тұтасады; сондықтан формадан мазмұн ізделеді, әдеби тілде мазмұнға арнап форма ізделеді;

г) жаңа лексемалар жасалымпаз; ситуацияға байланысты түсінікті бола беретін мағынасы қалай тартқыласаң да соған сәйкес келе беретін сөздер пайда болады. Олар әткге қатысушылардың барлығына таныс көптеген нақты тұрмыстық зат атауларының орнына қолданыла береді. Мысалы, мынау бір керемет дүние екен, әне бір затың ұнап отыр.

ғ) экспрессивті-бағалауыштық реңк күшті.

д) назар аударту, ден қойдырту мақсатында бір синтаксистік тұтастықта қайталап айтушылық басым;

е) сілтеу есімдіктерінің коммуникативтік қызметі басым;

ж) өз ойын өзі мақұлдайтын қызметтегі шылаулар, қыстырынды сөздер көп.

Адам өзінің МЕН-і мен МЕН-іне жақын айнала ішінде жеке тұлғасын қалыптастырады. Бейресми орта адамның шынайы болмысы мен табиғатын танытады. Бейресми ортада коммуникацияға түсетін тұлғалар туыстық, тұрмыстық, әріптестік, кейде қызметтестік контекст арқылы өзара тілсіз түсінушілігі басым болады. Адам өзіне ең жақын отбасы, бір үйде тұратын адамдарымен қарым-қатынас кезінде тіпті тілдік коммуникацияны өте аз қолданады, кейде тіпті қолданбайды деуге болады.

Ауызша тілдің корпусын жасаудағы басты мақсат қазақ тілінде дұрыс сөйлеу және қазақ тілінің сөзсаптамын зерттеуге материалдар жинау болатын болса, онда дәл осы кезеңдерде тұрмыстық қатынастағы қысқа-қысқа ауызекі диалогтарды емес және ресмиге жақын сипаттағы телесұхбат, пікірлер емес, дәріс, конференция, семинар, дөңгелек үстелдерде сөйленген сөз, монологтар емес, және бастысы қазір көп кездесетін қате сөйленген сөздерді емес, бейресми жағдайдағы қарапайым әңгімелерді, дәстүрлі әңгіме айту машығындағы сөздерді жинау дұрыс болады деп ойлаймыз. Ауызша сөйленген сөздердің барлығын корпусқа салып, мына сияқты түсініксіз мәтіндерден зерттеу нысанын табуға болмайды. Мысалы: Өткөнжол ол айтты/ жұмұшшұ тұуралы да//, тыңдадык/ иии сол ... аяқ кезінде[...] жаңағы йескертпелерден гейін//, содан гейін енді//, жаңағы қосұмша сөз айттық сынды//, қалай/ не жасадық деген мәселе//, енді ііііі ымммм кәзір біз енді///.[...]

Сөйлеу тілі дегенде жоғарыдағыдай «аузынан сөзі түскен» мәтіндер емес, тіл ұстарманың тұрмыстық санасын жарыққа шығаруға қызмет ететін, танытатын, түсінікті сөз стилін аламыз.

Сөйлеу тілінің бүгінгі қазақ ауызша коммуникациясындағы негізгі қызметтерін былай көрсетуге болады:

– коммуникативтік;

– коммуниканттардың шынайы тұлғасын танытады, МЕН-ін көрсетеді;

– ұлттық тілдің негізі;

– жағымды-жағымсыз, жақтыру-жақтырмау модальділігін, қарым-қатынас сипатын қалыптастырады;

– әдеби тілдің жетілуіне, өзгеруіне негіз болады;

– коммуникантардың әлеуметтік тең ортасын қалыптастырады;

– ауызша ресми тілде ресмилікті сейілту үшін пайдаланылады.

Ерекшеліктері мынада:

– құрылымдық сипаты күрделі;

– эмоционалды-экспрессивтік бояуы қанық, яғни эмоционалды-бағалауыштық, экспрессивті лексика басым болады;

– құрылымы басы артық сөздерден арылып отырады, аз тілдік құрылымға көп ақпарат сыйдыра алады, толық емес сөйлемдер, экспрессивті синтаксистік құрылымдар жиі қолданылады;

– Сұраулық және лепті сөйлемдер жиі қолданылады;

– Сөйлеу барысында, сол мезетте ойдан шығарылған жаңа сөздер пайда болады;

– Сөздер, бірліктер, фразалар мәтінге тәуелді болады, мәтіннен тыс түсіну мүмкін болмайтын қолданыстар көп кездеседі;

– Сөздің интонациясы, сөз просодиясы өте үлкен рөл атқарады.

Сөйлеу тілі – сөйлеушінің тұлғасына жақын, сөйлеушінің тұлғасын танытатын тіл болады, сондай-ақ сол ұлттың сөйлеу дәстүріне табан тірейтін, сөйлеу дәстүрін танытатын да сөйлеу стилі болып табылады. Сондықтан сөйлеу тілінің мәтіндері ұлттық корпусстың мазмұнын байытатын мәтіндер болады. Ұлттық сөзсаптам ерекшеліктерін көрсететін, ұлттық сөзсаптам үлгілерін беретін тіл болады.

Қазіргі өзге тілдің сөйлеу стилі араласқан, сөйлем құру, сөз тіркестіру амалдары еніп кеткен қазақ тілінің «бөгделенген» қолданысын түзеу үшін де, дәстүрлі сөзсаптам ерекшеліктерін зерттеу үшін де ауызша тіл корпусында дәстүрлі сөйлеу тілінің мәтіндері көптеп алынуы қажет.

Тіпті арнайы «дәстүрлі сөйлеу тілі» деген ішкорпустың болуы керек деп ойлаймыз. Дәстүрлі сөйлеу тілін үлгі етуді қойсақ, одан көз жазсақ, әдеби тілдің табиғаты өзгере бастайды.

Ауызша тілдің ішкорпусында, әсіресе ауыл қазақтарының ауызекі сөйлеу тілі, ақсақалдардың сөзсаптамы көптеп енгізілуі қажет. Ол үшін әлі де кеш болмай тұрғанда, мысалы, «Қазақ тілі дәстүрлі сөзсаптам үлгілерінің прагматикасы» атты мемлекеттік грант аясында бірыңғай қазақы ортасы бар елдімекендерге ұйымдастырылған іссапарлар сияқты, ғылыми экспедициялар жасалуы қажет.

Мысалы, сол іссапарлардың бірінде жазылып алынған Ақтөбе облысы, Шалқар ауданы, Айшуақ ауылының Қобылашев Әбдіраш деген мал дәрігері болып жұмыс істеген 80 жастағы ақсақалының екеуара әңгімесінен үзінді алып көрейік: Отырған о жерінде // орталыққа / бір ықпалың болуу герек /// тәрбійең болуу герек /// ақылыңнан тәрбійең болуу герек //жүргөн жүрүсүннен тәрбие болуу керек ///орта көргөндігің белгілі боп тұрұуы керек /// не босса ана/ана көпшіліктің көзүне айта бермеу керек/// әр әңгіме әркімге айтатын өлшем бар / сол өлшеммен айтұу керек ///көрдүңүз бә?/// ол әлгі мен сөз білед

екем деп/// не тыңдамайтын жерге сөздү сұхпау керек /// керег жоқ оған е-ең тийімдісі аузыңды баққан ол арада/// пәлеңшең бүй деп айтты дегізгеннен // Әдімандай балалы бол деген / Әдіман::: соғезде баланың аз кезі ед/// Әдіман оқутұушү//жеті балалы болған //сондұқтан: Әдімандай балалы бол дедім///жеті бала ғой енді///о гезде бала аз ///Әйнектей құу бол дегенім// құу адам ед// Карол дейтін ед //сақалы үлкөн // құу // өте құу адам ед // құу //өзү менің жездем ед/// Әйнектей құу бол///Әдімандай балалы бол/// Өтәлідәй у бол дейтінім оның тілі ашшы адам ед/// оның кәзір бір/екі мысалын айтам ///ол құу // ашшед // сондұқтан Өтәлідәй ұу бол//

Қайтадан тында/// Керімбайдай беріш бол істеген ісіне, қылуасына ыңғайлы// кімдей // Ілійәстәй ергіш бол, істеген ісіне қылуасына белгілі/// Әдімандай балалы бол / бала көп // Сыдықтай саналы бол тәрбийелі қазақ тілінің оқутұушысы тәлім/тәрбийесі жұрт/жұртшысы соны бөлөк Әйнектей ұу бол, ұу - отырған жері ///құу ол құу // құу // құу //Өтәлідәй ұу бол //ұу // ұу ол жәнәғі қайдағы пәлені айтад/// Мысал айтсам: *Қалбек дейтін қалық бірмен / бірі / отбасы отұрұп қалжыңдасад /// отбасы отұрұп қалжыңдасад /// бірімен / бірі /// ол аямайд // мінің гәрсе бірінің бірі::: бірәқ ұшуншү біреу гелсе//ол әңгіме қалад///шатасып отқанда да сырын байқатпайд /// Пұсұрман дейтін әкеміз/дің алты баласы болған///жұрт білмей не босса сон айтад/// алты баласы болған // жоқ Бес пұсұрман дейд/// / жо:қ Пұсұрман бесеу емес, алтау Қалбек//Қарабас//Өмір//Жарас//Темір/// көрдіңізба?! Білмейд///көпшілігі Өмір менен Темір аз үй болған // мүсәпірлеу болған сексен деген кісі/сексен деген кісі.*

Бұл – үйде, шай үстінде екеуара отырып, алдын ала дайындықсыз, машықсыз айтылған әңгіме болатын. Ақсақал әңгіменің таспаға жазылып жатқанын, бұл жазбаның белгілі бір мақсатта пайдаланылатынын сезіп отырды. Алайда сөздің тілдік сипаты емес, мазмұны, тарихи дерегі маңызды болады деген оймен, өзінің сөзсаптамын қадағаламады. Сол кезде керісінше, сөзі еркін, табиғи шықты.

Үш сағаттық сұхбаттасу аудиотаспадан өзгеріссіз жазбаға түсіріліп, мүмкіндігінше, фонетикалық транскрипциясымен берілді. Сұхбаттасқан уақытта сөз өте әсерлі, түсінікті, нақты болды. Оған бірден-бір әсер еткен сөйлеушінің интонация реңктерін шебер пайдалана білуі және мақсатты қайталаулар жасауы, сондай-ақ әр сөзін түсіндіріп отыруы болды деп ойлаймыз.

Сонымен бірге әдеби тілдің сөздіктерінде тіркелмеген тіркестер, жаңа мағыналар байқалды. Қазақ сөздік қорын байытатын, ұлттық сөйлеу тілінің белсенді қолданыстары шықты. Мысалы, орта көргөндік (тәрбиелілік, әдептілік, оқыған), отырған жер (елмен сөйлесіп, араласқан орта), сөз сұхпау (сөйлемей), аузыңды бағу (ауыз баққан деген біреудің сөзін аңду, өсек теру деген мағына бар еді, мұнда 2-жәкпен айтылып, аузыңды жауып отыру, үндемеу деген мағынада қолданылды), әңгіме өлшемі (сөз нормасы), мысалын айту (дәлел келтіру), қылуасы (әдеті), қайдағы бәле (небір қызық), халық (ру), отбасы отыру (отбасы мүшелерінен басқа ешкім жоқ кез), әңгіме қалу (тыныштала қалу, басқа әңгімеге көшу), шатасу (ұрсысу, дауласу, төбелесу), не болса, соны айту (аузына не келсе, соны сөйлеу), аз үй (аз ғана) деген сияқты сұхбаттың

жанын кіргізіп отырған көптеген тіркестер ана тіліміздің байлығы болар еді.

Бұрынғы уақытта адамдар өзіне ет жақын ортамен етене қатынаста болған. Әулет болып, бір ауылда тұрған. Сол кеңістіктен ұзай қоймаған. Үй-іші, әулет, туған-туыс, көрші-қолаң, құрбы-құрдас, ауыл адамдары сияқты әбден таныс ортада қазақтың дәстүрлі сөйлеу тілі басты коммуникативтік құрал болды. Алайда заман өзгеріп, тума-туыс бірыңғай отыратын қазақы ауылдың ішіне өзге рудың, бөтен ұлттың адамдары «келіп кіріскен» (А.Байтұрсынұлының сөзі) заманда, қоғамдық формация ауысып, әлеуметтік-тұрмыстық, еңбек қатынастары өзгерген тұста дәстүрлі ауызекі сөйлеу тіліне де өзгерістер ене бастады. Таныс орта енді таныс-бейтаныс ортаға айналып, тілдік жағдаят басқашаланды.

Сөйтіп, ауызекі сөйлеу тілінен сөйлеу тілі немесе бейтарап сөйлеу тілі қалыптасты. Бұл қарияның тілінен де бірыңғай ауызекі сөйлеу тілі емес, бейтарап сөйлеу тілі, әдеби тілге жақындау сөйлеу тілі байқалды.

Тағы да Қобылашев Әбдіраш ақсақалды бір тыңдайық: Бұрұн / әрнәрсе өзүнүң орнұнда болатын. Әрнәрсе. Қайнаға / қайнаға гүйүнде // келін, келін гүйүнде // атене / атене гүйүнде болатын. Кәзір онұң бәрі бостаң гетіб баратыр, қарағым. Кәзір /// бұрұн атана бірінші орұнда болатын. Сол үйдің башшысы / сол үйдүң ақылшышы болатын / сонұ тындайтын. Сосен /// бұл дұрысқой /// атаның бәрі білгір / бәрі әқылды деп айта алмаймын / бәрі де білет / бірәқ жастар өзү тындауын тындайтын да / өзүнүң деңгейіне ғарайтын. Кәзір олай емес/ себебі / ала / екі түрлү /// тындайтындар бар // тындамайтындар бар. Осұған не герек дейтіндер бар. Қәзіргі атаналар басым бағыб кетті, басым бағұуға гөштү / өйткөнү кезінде / бала ғоштайт, бала ғоштайм деп те ғошталмайт шығар отбасының бүтүншүлігі герек. Кәзір бағыт бағдар өзгөрүп кетті, айналайын. Қыйын бүрөуге бүрөу қожалық // бүрөуге бүрөу ақыл-генес сұрамайт, өзүм білем гөбөйіп кетті.

Бұл 2 минутқа жетпейтін ауызша сөзде қазіргі қазақ тілі сөздіктеріне енебеген бірнеше сөз тіркесі мен жаңа мағыналар бар. Мысалы: әрнәрсе (әр зат), бостан кету (босаңсып кету), басын бағып кету (өз басын ойлау, басымен қайғы болу), отбасының бүтіншілігі (отбасының тыныштығы, отбасы болып, шашауы шықпай, бір үйдегі тату тіршілік), біреуге қожалық ету (біреуге айқыл айту, біреуді билеу).

Мақсатты қайталаулардың грамматикалық та, лексикалық та үлгілері бар (болатын, -атын формасының бір сөз ағынында қайталануы; әрнәрсе сөзінің аздаған кідіріспен қайталануы; қарияның жасқа қарата айтатын ұлттық қаратпа атауыштары бірін бірі қайталамай, 2 түрлі тұлғада, бірдей арақашықтықта айтылды: 1) қарағым, 2) айналайын) сөйлемдер қазақ тілінің синтаксистік нормасына сай мәтін үшін аяқталған болды.

Экспедиция арқылы материалдар жинап, оны аудионұсқасымен бірге жазба нұсқаға түсіріп, ауызша корпусқа енгізіп, зерттеуге де, қазақ тілінде дұрыс сөйлеуге де үйрететін мазмұнды қор жинауға болады.

«Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту Интернет-ресур-

сы» атты мақсатты қаржыландырылатын тақырып аясында ауызша тілдің іш-корпусын жасау да жоспарланып, жұмыс басталып кетті. Ауызша тілдің іш-корпусына: 1) күнделікті тұрмыстық қатынастағы ауызекі сөйлеу стилі (үй-іш, отбасы, туған-туыс, көршілер арасындағы диалогтар, дүкен, көлік, қоғамдық орындардағы тілдесулер, бейресми сипаттағы телефон арқылы сөйлесулер), 2) бейресми әңгіме, сұхбаттардағы машықсыз, дәстүрлі сөйлеулер, 3) ауызша шаршысөз, дәріс, конференция, дөңгелек үстел, кеңес, семинар, сұхбаттардағы алдын ала жоспарланған тақырыпқа сөйленген сөздер, 4) кино, сахналық қойылымдағы дискурстар, 5) ауызша сөйленген машықты, машықсыз діни дискурстар, 6) теле-радиодағы алдын ала тақырыбы белгілі бағдарламалардағы машықсыз сөздер алынады. Ал жазба мәтінді дыбыстау, өлең, қара сөзді жатқа оқу, ақпараттық хабарлардағы жазылған мәтінді оқып беретін ауызша сөздер бұл корпусқа әзірге алынбайды. Ал аталған тақырып бойынша 1) бейресми әңгіме, сұхбаттардағы машықсыз, дәстүрлі сөйлеулер, 2) ауызша шаршысөз, дәріс, конференция, дөңгелек үстел, кеңес, семинар, сұхбаттардағы алдын ала жоспарланған тақырыпқа сөйленген сөздер жиналып, транскрипциясы жасалып, енгізілді.

Зерттеушілердің компьютерлік тілдік қорды «ғылым адамының өз зерттеу нысанына жаңаша тұрғыда көз салу мүмкіндігі. ... тілдік қор неғұрлым қомақты болса, солғұрлым тіл құрылысының сыры тереңірек ашылады, зерттелетін нысан жөніндегі түсініктердің аумағы кеңейеді, зерттеушінің мүмкіндік-қабілеті әлденеше есе артады, шығармашылық қуат көздері ашыла түседі, сөйтіп, бұл жаңа мүмкіндіктер қазақ тілінің жүйесін жетілдіруге және тіл жүйесін мұқият тануға жұмсалатыны сөзсіз» [12, 7 б.] дегеніндей, ауызша тіл корпусының беретін сандық және сапалық ғылыми ақпараты көп болады деген ойдамыз.

2.3 Айтылым үдерісіне қатысты ауызша әдеби тілдің нормалары

Қазақ тілінің ең жоғарғы ұлттық формасы – әдеби тіл. Әдеби тіл нормалары сұрыпталу, қалыптандыру сияқты үдерістерден үнемі өтіп отырады. Себебі, тіл – жанды құбылыс. Тіл – қоғаммен бірге өмір сүреді. Қолданушылардың тілді меңгеру және қолдану деңгейіне байланысты ұлттық әдеби тілдің дәстүрлі нормалары сақталады не жойылады, дамуындағы негізгі үрдістер айқындалады не айқындалмайды, сөздің коммуникативтік сапасы артады не кемиді. Сондықтан әдеби тілдің кодификацияланған нормаларын үнемі назарда ұстап, оның қалыптануына барлық жағдай жасалуы тиіс.

Қазіргі кезде әдеби тіл нормаларының заңдылықтары айқындалып тұрғанымен, оны тиімді қолдану жағы әлі де ақсап тұр. Ғалымдар қазіргі қазақ тіліндегі айтылым нормалары жазылым нормаларына «сәйкес» немесе ауызша сөз нормаларының «жойылу» тенденциясы басым екенін айтып дабыл қағып отыр [29].

Ауызша сөздің жалпы тіл мәдениетіне қойылатын талаптарымен қатар өзіне тән нормалары бар. Оның нормаларына сөйлеу барысында сөздерді дұрыс дыбыстау, сөздің сазын келтіріп айту тәртібі жатады. Сөйлеген сөздің әсерлі, құлаққа жағымды, көңілге қонымды болуы, оның сазды айтылуынан болса,

сөздің әсерлі-әсерсіздігі адамның эмоционалды-экспрессивтілігіне, көңіл-күйіне байланысты болуы қазақ тіліндегі орфоэпиялық және интонологиялық заңдылықтармен өлшенеді.

Қазақ тілінде орфоэпиялық нормалардың жүйелі түрде заңдылықтарын белгілеген, кодификациялық нормасын саралаған ғалымдардың бірі – Р.Сыздық. Р.Сыздықтың «Сөз сазы» еңбегінде алғаш рет ауызша сөз нормаларын 36 параграфке топтастырып береді [30].

Кодификацияланған ауызша сөз нормасы дыбыстау деңгейімен, лексикалық, грамматикалық және стилистикалық деңгейлермен өлшенеді. Дыбыстау деңгейінде – сегментті (фонетика-фонология) және суперсегментті (интонология, акцентология) бірліктерді жөнімен қолдану жатады. Ал ауызша тілдің лексикалық деңгейі сөйлесім үдерісіне түсетін коммуниканттың сөздік қорына байланысты. Ауызша әдеби тілдің лексикалық нормасы мыналарды қамтиды: а) әр сөздің мағыналық реңктері бойынша қолдану. Кез келген сөздің (бейтарап сөздер басқа) реңктік сипаты болады. Айталық, адамға немесе жануарларға байланысты айтылатын сөз, деректі не дерексіз затқа байланысты айтылатын сөз, жалпы не жалқылықты білдіретін сөз т.б. Мысалы, мекендейді сөзі жануарға, ұлт немесе этносқа қатысты айтылады. Ал жеке адамға қатысты бұл сөз қолданылмайды; ә) мәндес (синоним) сөздерді орнымен қолдану (тәртібін, қолайын, реттілігін ескеру). Мысалы, адамның жасына байланысты айтылатын жасар, жастағы, жасқа келген, жасаған деген синонимдік қатарлардың қолданымдық нормасы әрбасқа. Адам жасының бірден жиырмаға дейінгі кезеңін сипаттағанда бір жасар, он екі жасар, жиырма жасар деуге болады, ал қырық жасар, сексен жасар деп қолдану – нормадан жөнсіз ауытқу. Отыз бен елу-алпыс жас аралығындағы адамдарды елу жастағы, елуге келген деп, жас шамасы едәуір ұлғайғанда, тоқсан жасаған, жетпіске келген деп айтқан дұрыс деп саналады; б) тұрақты тіркестердің құрамдық компоненттерін (сөздерін) қысқартпай не өзге сөзбен ауыстырмай, мәтіндегі орнын тап басып қолдану; в) термин сөзді өз мағынасы мен дефинициясына сәйкес қолдану; г) көне, сирек қолданылатын сөздердің мағынасын анық білмей қолданбау; ғ) сөздің денотаттық, коннататтық, сигнификативтік мағынасына сәйкес қолдану т.б. Грамматикалық деңгей бойынша грамматикалық категорияларды, синтаксистік конструкцияларды нормаға сәйкес қолдану. Синтаксис бойынша – ауызша және жазбаша сөздің жүйесіне сәйкес синтаксистік құрылымдарды және сөздердің байланысу амалдарын дұрыс қолдану. Морфология бойынша – жалғау мен жұрнақтардың негізгі сөзге беретін реңктік мағынасын ескеру т.б. Стилистикалық деңгей бойынша сөз дұрыстығы (әдеби тіл нормасын білу), сөз түсініктілігі (сөзді адресатты ойлап тандау, бейәдеби лексиканы ажырата алу), сөз тазалығы (бөтен сөзден қашу, варваризмдер, кальканы білу), сөз дәлдігі (сөздердің тіркесім валенттілігін, семалар өрісін білу), сөз бедері (стильдік жағдаятқа тән қолданыстарды, стильдік маркерлі бірліктерді ажырату), сөз байлығы (гиперонимдік, гипонимдік қатынастағы лексикалық бірліктерді ажырату, грамматикалық форма алуан түрлілігінің функционалдық қызметін білу).

Әр тілдің орфоэпиясы сол тілдің басты фонетикалық заңдылықтарына сүйенеді. Мысалы, орыс тілінде сөздің дұрыс айтылуы екпінге байланысты

болса, қазақ тілінде үндесім заңдылығына табан тірейді. Қазақ тілінің дыбыс жүйесінде әсіресе дыбыс үндестігі ерекше орын алады. Сөздерді жазылуынша айту әсіресе жастардың тілінде үйреншікті машыққа айналып бара жатқанын аңғару қиын емес. Мұны тіліміздің табиғи дамуының нәтижесі деп қарауға болмайды. Қазақ тілінің орфоэпиялық нормасы және оның маңызы туралы алғаш пікір білдірген ғалымдардың бірі профессор М.Балақаев өзінің «Қазақ тілі мәдениетінің мәселелері» деген еңбегінің алғы сөзінде былай дейді: «...Бірақ сөздердің айтылу нормасы және сол норманы сақтап сөйлеу мәселесі ескерілмей, елеусіз қалып келеді. Сондықтан да бізде бұл мәселе жөнінде күні бүгінге дейін ала-құлалық байқалады. Мына сөз былай айтылмайды, былай айтылады, мынасы дұрыс, мынасы теріс дегенді көп есітпейміз. Кім қай сөзді қалай айтады, қалай есітеді-бәрібір әйтеуір нобайы келсе болғаны. Бұл кемшіліктерден арылу үшін әрбір сөздің дұрыс айтылу нормасын сақтап сөйлеуді тіл мәдениетінің басты талаптарының бірі етіп қою керек» дей отырып, сөздердің айтылу нормасы құлаққа жағымды, көңілге қонымды айтылғанның бәрі – норма, бәрі дұрыс. Құлаққа жат не тосаң естілетін, көңілге қонбайтын бұлдыр, оғаш түсініксіз, дөрекі айтылғанның бәрі нормаға жатпайды, дұрыс емес» дейді [31]. Шындығында да, ғалымның сөзіндегі құлаққа жағымды, әуезді дегеннің ауызша сөз үшін маңызды. Себебі тілдің табиғи заңдылығы сақталса, сөз әуезді әрі әуенді шығады.

2.4 Қазақ тілінің ауызша корпусын құрудың ғылыми-практикалық мәселелері

Әлемдік ұлттық корпус құру тәжірибесіне көз жүгіртетін болсақ, корпус-тың қажетті құрамдас бөлігі ретінде тек жазбаша ғана емес, ауызша сөйлеу формасымен бірлікте қарастырылатынын көрсетеді. Тілдің ауызша формасынсыз корпус ұлттық тіл туралы ақпараты толық деп саналмайды. Сондықтан көптеген елдерде, оның ішінде ағылшын тілінің ауызша корпусы (Spoken English Corpus) [32], француз тілінің ауызша корпусы (corpus de la Parole) [33], қытай тілінің ауызша корпусы (Chinese Spoken Language Corpus) [34], испан тілінің ауызша корпусы (Corpus del Español Oral) [35], жапон тілінің ауызша корпусы (Japanese Spoken Language Corpus) [36], түрік тілінің ауызша корпусы (Turkish Spoken Language Corpus) [37], чех (Český národní korpus / ČNK) [38] және орыс тілінің ауызша корпусы [39] жасалған.

Чех, ағылшын, эстон, француз тілінің ауызша корпусында ауызша мәтіндер тақырыптық, жанрлық ерекшеліктерге байланысты жеке-жеке топтастырылғанын көреміз. Айта кетерлік жайт, неміс тілінде тек ауызша корпусқа байланысты 70-тен аса сайт бар. Олар өз ішінде дайындықты, дайындықсыз, академиялық, іскери, ауызекі ауызша сөзге, сондай-ақ ауызша диалектілік, просодикалық ерекшелікке байланысты, радио, телевидение, интернет кеңістігіндегі ауызша сөздер жеке-жеке топтастырылған.

Ауызша корпустар туралы мәлімет

Тіл	Корпус	Сипаттамасы
Чех	V1 диалектісі: көп деңгейлі транскрипциясы бар диалектілік корпусы	дәстүрлі диалектологиялық материал және монологтік типтегі спонтанды сөздер, Көлемі: 100 000 сөз.
	ORAL 2013: чех тіліндегі бейресми сөздер корпусы (транскрипция және аудио)	Бұл корпуста бейресми сипаттағы әңгімелер топтастырылған. Көлемі: 2,8 миллион сөз.
	Prague DaTabase of Spoken Czech 1.0	Спонтанды диалогтар топтастырылған. Көлемі: 770 000 сөз.
Ағылшын	Boston University Radio Speech Corpus	Бұл корпуста радио жаңалықтарының жазбалары мен мәтіндері жинақталған. Көлемі: 7 сағат.
	ELFA	дәрістер мен семинарлардың жазбалары топтастырылған. Көлемі: 13 сағат.
	Busckeye ауызекі сөйлеу корпусы	Сұхбаттар корпусы Көлемі: 5 сағат.
	BNC2014 ауызша корпусы	Бұл корпуста Британдық ағылшын тілі ана тілі болып табылатын адамдар арасындағы ауызша сөздер. Көлемі: 10 миллион сөз.
Эстон	Дәрістер корпусы	Бұл корпуста конференцияларда сөйленген сөздер, әртүрлі академиялық дәрістер мен ауызша баяндамалардың жазбаларынан тұрады. Көлемі: 41 сағат.
	Радиосұхбат корпусы	Әртүрлі радиобағдарламалардан және телефон сұхбаттарынан алынған сөздер топтастырылған. Көлемі: 36 сағат.
	Радиожаңалықтар корпусы	Бұл корпуста хабар тарату жаңалықтары топтастырылған. Көлемі: 19 сағат.
	Диалектілік корпус	Сұхбаттар топтамасы. Көлемі: 1,3 миллион сөз.
	Эмоционалды сөйлеу корпусы	Бұл корпуста ашулану, қуаныш, қайғы сияқты эмоционалды сөйлеу мәтіндері топтастырылған. Көлемі: 1234 сөйлем.
	Дайындықсыз сөйлеу корпусы	Бұл корпуста әртүрлі диалектілік ерекшеліктер жиналған. Дайындықсыз сөйленген сөздер топтастырылған. Көлемі: 635 000 сөз, 90 сағат.
Француз	CLAPI	Бұл әртүрлі контексттердегі әлеуметтік өзара әрекеттесуді сипаттайтын 40-қа жуық корпусты қамтитын жинақ: кәсіби, жеке, институционалдық, коммерциялық, медициналық және білім беру бойынша топтастырылған. Көлемі: 323 595 сөз.

Corpus de Français Parlé Parisien des années, 2000 г.	Бұл корпус сұхбаттардан тұрады
Корпус изучения современного французского языка	Бұл корпуста пікірталас, әдеби және ғылыми мәтіндер, аймақтық және ұлттық арналардағы ауызша сөздер Көлемі: 10 миллион сөз, 350 сағат

Түрік тілінің ауызша корпусында 1,5 миллионнан астам сөз және 200 мыңға жуық сөз формалары қамтылса, британдық ұлттық корпустың (BNC) 10%-ы ауызша сөйлеу жазбаларынан тұрады. Орыс тілінің Ұлттық корпусы тәжірибесін айтатын болсақ, орыс тілінің Ұлттық корпусында диалектісі бар ауызша сөз түрлері және әдеби ауызша сөз материалдары енгізіледі. Сондай-ақ орыс ғалымдары ауызша сөз корпусын жасауда оның көлемінен бұрын жанрлық және тақырыптық тепе-теңдікті сақтауға және ауызша корпусты 10 миллион сөзқолданысқа жеткізуді көздеп, ауызша корпусты аудио жазбалармен, транскрипциялық жолмен және көбінесе ауызша сөзді орфоэпиялық емес, орфографиялық стандартын сақтауды жөн көреді. Орыс тілі корпусында ауызша сөз орыс тілінің дәстүрлі орфографиясы бойынша жазылады деп шешіледі. Оны *щас, тыща* сияқты мысалдардан көруге болады. Ал сөйлем ішіндегі пунктуациялық белгілер мүлде ескерілмейді немесе слэштермен, яғни қиғаш сызықтармен алмастырылып беріледі. Мәтінде нүкте қызметін атқаратын тыныс белгілері ғана сақталады. Осы ретте, қиғаш сызықтың ешқандай мағыналық жүгі жоқ екендігін (яғни паузаға, т.б. сәйкестендіру міндеті емес), тек оқуға ыңғайлы болу үшін қойылатынын да байқауға болады. Мысалы: [Андрей Себрант, муж, учёный] Я сам помню пару лет назад наблюдал/ и это меня сильно впечатлило/ катинку/ как ещё не умеющий говорить реально малыш там двух/ что ли/ лет добирается до планшета/ совершенно осознанно находит на нём иконку Скайп и запускает Скайп/ там находит бабушку по её иконке/ запускает сеанс связи с бабушкой и начинает ей агукать и улыбаться.

Орыс тілі корпустағы қиғаш сызықтар (слэштер //, /) фонетикадағы синтагмалық бөлшектерді бөлуге пайдаланылатын шартты белгіге тұрпаты ұқсас болғанымен, мазмұны корпуста сақталмаған. Яғни бір қиғаш сызық – кіші кідірісті, ал қос қиғаш – үлкен кідірісті білдірмейді. Сызықтар (слэштер) тек оқуға түсінікті болу үшін қойылған. Демек, қазақ тілінің ауызша ішкі корпусын жасауда ауызша сөздің типтері мен түрлерін, жанрларын анықтау ең бірінші мәселе екені көзге түседі, екінші мәселе жазбашаға түскен сөздің орфоэпиялық норма бойынша ма, әлде орыс тіліндегідей орфографиялық стандарт формасымен жазу мәселесін шешу керектігі байқалды. Үшінші мәселе, мәтіндердің синтагмалық топқа сөйлеушінің сөзіне бағдарлап қою немесе мәтіннің оқырманға түсінікті болу жағы қарастырылды.

Ауызша корпусты талдау келесі қадамдардан тұрады:

1. Транскрипция: ауызша мәтіндердің жазбалары талдануы үшін жазбаша сөзге түсірілуі керек. Транскрипция сөзбе-сөз (сөздер мен сөз тіркестерін дәл қайта жазу) немесе фонетикалық (дыбыстар мен сөйлеу интонациясын бейнелеу) болып екіге бөлінеді. Қазақ тілінің ауызша ішкорпусында аталған

екі транскрипция беріледі. Сонымен қатар орфоэпиялық норма да көрсетіледі.

Сұхбат беруші	Орфография	Сөйлеуші орфоэпиясы	Бейнефильм	Орфоэпиялық норма
Асанәлі Өшмөв	Кино – ең бірінші өнер. Сол кино өнерін дамытуда үлкен үлес қосқан азаматтар бар. Баршылық. Соған қуанам. Бұрын аз болатын, бірақ маз болатын, қазақша айтқанда. Шәкең бастаған ұлы топ біраз дүние жасады. Солардың бәрін алға қарай апарушы – әзіргі жастар. Соларға сенем. Сұлтанмахмұт Торайғыров айтпақшы, «мен жастарға сенемін». Талай фестивальға қатысқан, фестивальдарда жүлде алған талай шығармалары бар.	↑кино – / → ^{ең} бірінші ^ө нөр ^і /// ↑сол кино ^ө нөр ^{ін} → // алға дамытуда [↑] / үлкен үлес қосқан → // азаматтар бар [↓] /// ↑баршылық [↓] /// ↑соған қуанам [↓] /// ↑бұрын / аз болатын [↓] /// ↑бірақ маз болатын [↓] // қазақша айтқанда [↓] /// ↑шәкең бастаған → / ұлы топ // біраз дүние жасады [↓] /// ↑солардың бәрін → / алға қарай / апарушы → // қазіргі жастар [↓] /// ↑соларға сенем [↓] /// ↑сұлтанмахмұт торайғыров айтпақшы → // «мен жастарға сенем ^і » [↓] /// ↑талай фестивальға қатысқан → // фестивальдардан жүлде алған → // шығармалары бар [↓] ///	▶	↑кйно – / → ^{ең} бірінші ^ө нөр ^і /// ↑сол ^г йно ^ө нөр ^{үн} → // алға дамытуда [↑] / үлкен үлес қосқан → // азаматтар ^б ар ^і /// ↑баршылық [↓] /// ↑соған ^ғ уанамың [↓] /// ↑бұрын / аз болатын → бірақ / маз ^б олатын [↓] қазақша айтқанда [↓] /// ↑шәкең бастаған → / ұлы топ // біраз дүние жасады [↓] /// ↑солардың бәрін → / алға қарай апарушы → // қазіргі жастар [↓] /// ↑соларға сенемін ^і /// ↑сұлтанмахмұт торайғыров айтпақшы → // «мен жастарға сенемін» [↓] /// ↑талай фестивальға қатысқан → // фестивальдардан жүлд ^д алған // → шығармалары ^б ар ^і ///

Сурет 1. Ауызша корпустағы просодикалық белгіленім конкордансы

Екінші бағандағы сөйлеушінің орфоэпиясында сөйленген сөз орфоэпиялық транскрипция арқылы беріледі. Яғни, сөйлеушінің өзіндік орфоэпиялық нормасы жазылады. Ал төртінші бағанға сөйлеушінің орфоэпиялық транскрипциясы емес, орфоэпиялық норма бойынша беріледі. Мысалы, ↑кино – / →^{йең} бірінші ^өнөр^і[↓] /// ↑сол кино ^өнөр^{ін} → // алға дамытуда[↑] / үлкен үлес қосқан → // азаматтар бар[↓] /// Бұл мәтіннен сөйлеушінің өзіндік орфоэпиясын көруге болады. Ал осы қолданысты нормаға түсірілген бағанды салыстырып қарастыратын болсақ, бірнеше айырмашылықты көруге болады: ↑кйно – / →^{йең} бірінші ^өнөр^і[↓] /// ↑сол ^гйно ^өнөр^{үн} → // алға дамытуда[↑] / үлкен үлес қосқан → // азаматтар ^бар^і[↓] ///. Авторлық орфоэпияда кино, үлес сияқты сөздердің орфоэпиялық нормаға сәйкес еместігін байқауға болады. Қазіргі сөйлеушілердің тілінде бұл құбылыс жиі байқалады [9].

2. Сегментке бөлу: ауызша мәтіндерді талдау үшін жазба сөзге түсіріліп, жазба сөздің тілдік бірліктері болып табылатын пунктографтар қойылды. Бұл әлемдік деңгейде белгілі басқа да корпустарда кеңінен қолданылатын әдіс. Яғни ауызша мәтінді дұрыс түсіну үшін пунктуациялық белгілер қойылды. Мысалы:

Кино – ең бірінші өнер. Сол кино өнерін дамытуда үлкен үлес қосқан азаматтар бар. Баршылық. Соған қуанам. Бұрын аз болатын, бірақ маз болатын, қазақша айтқанда. Шәкең бастаған ұлы топ біраз дүние жасады. Солардың бәрін алға қарай апарушы – әзіргі жастар. Соларға сенем. Сұлтанмахмұт Торайғыров айтпақшы, «мен жастарға сенемін». Талай фестивальға қатысқан, фестивальдарда жүлде алған талай шығармалары бар [40].

Көріп отырғанымыздай, сөйлеуші сөзі жазбаша тілдің заңдылығы бойынша берілген. Сөйлеушінің сөзіндегі просодикалық белгілер жазба тілдің заңдылығы түсірілген. Ал ауызша ішкорпустың «сөйлеушінің орфоэпиясы» атты бөлігінде сөйлеушінің барлық сөзі дәлме-дәл беріледі. Сөйлеушінің интонациялық, синтаксистік конструкциясын шартты таңбалар арқылы көруге болады. Мысалы:

↑кино / →^{йең} бірінші ^өнөр^і[↓] /// ↑сол кино ^өнөр^{ін} → // алға дамытуда[↑] / үлкен үлес қосқан → // азаматтар бар[↓] /// ↑баршылық[↓] /// ↑соған қуанам[↓] /// ↑бұрын / аз болатын[↓] // ↑бірақ маз болатын[↓] // қазақша айтқанда[↓] ///

↑шәкең бастаған →/ ұлы топ // біраз дүние жасады↓/// ↑солардың бәрін →/
алға ғарай/анарушы →//қазіргі жастар↓///↑соларға сенем↓///↑сұлтанмахмұт
торайғыров айтпақшы →// «мен жастарға сенем»↓/// ↑талай фестивалға
затысқан →// фестивалдардан жүзде алған //→ шығармалары бар↓/// [40]

Корпуста сөздің «дыбыстық» компонентісіз зерттеушіні фонетикалық деңгейде ауызша мәтінмен жұмыс істеу мүмкіндігін күрделендіреді. Ауызша сөзде сегментті бөлшектермен қоса, суперсегментті бөлшектер, сонымен қатар дыбыстық немесе транскрипциялық түпнұсқаға жүгіну мүмкіндігінің болмауы аудио материалдарды декодтау кезінде пайда болатын екі мағыналы немесе түсініксіз аудио материалдардың туындауына әкеп соғады. Сонымен қатар, стенограмма тілдің барлық басқа деңгейлерімен – морфология және сөзжасам, синтаксис, семантикамен, сондай-ақ ауызша сөйлеудің риторикалық аспектілерімен тиімді жұмыс істеуге мүмкіндік береді. Сондықтан ауызша корпуста сегментті бірліктермен қатар суперсегментті бірліктер де көрсетілді.

Ауызша корпуста аудиторлық (субъективтік) талдаудың мақсаты – талдау нысаны ретінде алынған дискурстың интонациялық, орфоэпиялық транскрипциясын мүмкіндігінше дәл берумен байланысты. Жоғарыдағы мысалдан көріп отырғандарыңыздай, бұл жерде жазба тілдің нормалары сақталмайды, яғни пунктографтар және бас әріп нұсқасы қойылмайды.

Ырғақты топтардың ара жігін көрсететін, синтагмааралық бөліну сызығы ретінде (/) алынған. Арасында осы таңба тұрған сөз бөліктері кідіріссіз айтылады, дегенмен олардың ара жігінде әуен өзгерісі болуы ықтимал дегенді білдіреді. Ал екі немесе үш (//), (///) сызығы синтагмалар ара жігі мен синтагмааралық кідірістің таңбасы. Арасында осы таңба тұрған сөз бөліктері бір-бірінен дауыс кідірісі арқылы бөлініп, әр қайсысы жеке-жеке әуен өрнегімен айтылатындығын көрсетеді.

Қазақ тілінде сөйлеу барысындағы интонацияның әуен компонентінің қызметіне сәйкес негізгі тон жиілігінің түрлі бағыттары бар екені белгілі. Олар: көтеріңкі, бәсең, бірқалыпты, көтеріңкі – бәсең, бәсең – көтеріңкі. Осыған сәйкес сөйлеушінің сөзіндегі негізгі тон жиілігінің қозғалыс бағыттары көтеріңкі болса – ↑, бәсең – ↓, бірқалыпты – → таңбалары алынды. Бұл талдау сөйлемнің жалпы коммуникативтік түрлерін анықтау (сөйлемнің жалпы мағынасы мен оның хабарлы, сұраулы, бұйрықты, лепті түрлерін айқындау) үшін, сөйлем, мәтін мазмұнының өзгешеліктерін, ерекшеліктерін ажырату үшін, мәтіндегі сөйлемдер мен синтагмаларды бір-бірінен ажырату, бөліп алу немесе басын қосу (мәтіндегі сөйлемдердің мағынасына қарай бір-бірінен жігін ажырату, сөйлем түрлерін мәнерлеп айту) үшін, эмоциялардың түрлерін айқындау (сөйлемнің коммуникативтік мағынасымен бірге әр түрлі эмоциялық-экспрессивтік түрлерін айқындайды, адамның көңіл – күйі (эмоциясы), сөйлемнің негізгі мағынасының өзгеруі) үшін қойылды.

Ауызша корпуста әрбір фонетикалық бірлік туралы толық ақпарат алуға болады. Корпуста дыбыстың артикуляция-акустикалық негіздемесі беріледі, яғни дыбыстың жасалымы, айтылымы, естілімі туралы ақпарат толық беріліп, сөздің жасалуына негіз болып тұрған просодикалық ерекшелігі де талданады. Оны төмендегі кестелі суреттен көруге болады.

өкемнің			
Фонетика-фонология			
	Жасалым	Айтылым	Естілім
ө	ашық, езу, тіл ортасы дауысты дыбыс	ауыз қуысты, кең шығысты дауысты дыбыс	жіңішке, ашық езу өуезді дауысты дыбыс
к	тіл ортасы, тоғысыңқы, тербеліссіз (қатаң) дауыссыз дыбыс	нық тоғысыңқы, ауыз қуысты дауыссыз дыбыс	күшті лепті, қатаң (шуыл), жіңішке өуезді дауыссыз дыбыс
е	емеурін, тіл ортасы дауысты дыбыс.	ауыз қуысты, емеурін шығысты дауысты дыбыс	жіңішке дауысты дыбыс
м	ерін (ерін-ерінді), тоғысыңқы, тербелімді (үнді) дауыссыз дыбыс	бос тоғысыңқы, көмей қуысты дауыссыз дыбыс	бос лепті, үнді (ыңыл), жуан өуезді дауыссыз дыбыс
н	тіл ұшы, тоғысыңқы , тербелімді (үнді) дауыссыз дыбыс	бос тоғысыңқы, көмей қуысты дауыссыз дыбыс	бос лепті, үнді (ыңыл), жуан өуезді дауыссыз дыбыс
і	қысаң, езу, тіл ортасы дауысты дыбыс	ауыз қуысты, қысаң шығысты дауысты дыбыс	жіңішке, қысаң езу өуезді дауысты дыбыс
ң	тілшік тоғысыңқы, тербелімді (үнді) дауыссыз дыбыс	бос тоғысыңқы, көмей қуысты дауыссыз дыбыс	бос лепті, үнді, жуан өуезді дауыссыз дыбыс
Просодика			
	Орфоэпия	Тембр	Буын
1		Жіңішке өуезді	ө-кем-нің

Сурет 2. Ауызша корпустағы фонетика-фонологиялық белгіленім терезесі

3. Жіктелуі: Ауызша мәтіндер әртүрлі критерийлер бойынша жіктеледі. Мысалы, тақырыбы (актерлік өнер, өмірбаян, әдебиет, өнер, тіл білімі), сөз стилі (мысалы, көркемсөз, шаршысөз т.б.), сөйлеушінің әлеуметтік деңгейі (екітілді, көптілді), сөйлеушінің сөзсаптам деңгейі (жоғары, орта, ортадан жоғары, төмен) және т. б.

Ауызша корпустың негізгі параметрлеріне мыналарды жатқызуға болады:

1. **Ауызша және ауызекі тілдегі сөз:** ауызша корпуста диалог, монологтар, сұхбаттар, пікірталастар және т.б. сияқты ауызша және ауызекі тілдің әртүрлі жанрлары негізге алынады.

2. **Тілдік формалардың әртүрлілігі:** ауызша корпуста әртүрлі диалектілер, жаргондар және басқа тілдік құбылыстар болады. Ауызша тілдің ерекшелігіне сәйкес редукция, протеза, эпитеза сияқты тілдік құбылыстар жиі қолданылады.

3. **Тақырыптардың әртүрлілігі:** ауызша корпус күнделікті әңгімелерден бастап мамандандырылған білім, ғылым салаларына дейінгі әртүрлі тақырыптарды қамтиды.

4. **Сөйлеушілердің әртүрлілігі:** ауызша корпус жас, жыныс, білім, кәсіп және т.б. сияқты әртүрлі әлеуметтік топтағы тілиелмендердің сөзі негізге алынады.

5. **Контексттердің әртүрлілігі:** ауызша корпус әртүрлі сөйлеу жағдайларын қамтуы мүмкін, мысалы, үй жағдайында, білім беру мекемелерінде, жұмыста және т.б. Осыған байланысты сөз қолдану мәдениеті де әртүрлі болады.

Ауызша корпус басқа корпус түрлерінен ерекшеленеді. Оларды атап айтамын болсақ:

1. **Форматы:** ауызша корпус – бұл мәтіндерден тұратын жазбаша корпусқа қарағанда ауызша сөйлеу жазбалары. Жазбалар аудио немесе бейне тасымалдағыштар арқылы жасалады.

2. **Табиғилық:** ауызша корпус ана тілінде сөйлейтіндердің табиғи ауызша сөйлеуін, оның ішінде интонация, кідірістер және т.б. сияқты барлық ерекшеліктерін көрсетеді.

3. **Стихиялылық:** ауызша корпустың ауызша сөйлеу көбінесе стихиялы, дайындықты емес, спонтанды жүзеге асады. Жазбаша мәтіннен айырмашылығы да осында.

4. **Вариативтілік:** ауызша корпус тілдің әртүрлі нұсқаларын, соның ішінде диалектілерді, әлеуметтік топ жіктерінің сөздерін және сөздің әртүрлі стилдік ерекшелігін көрсете алады. Бұл әсіресе уақыт өте келе тілдің өзгеруі мен тілдің өзгеруін зерттеуде пайдалы бола алады.

5. **Өңдеудің күрделілігі:** ауызша корпус жазбаша корпусқа қарағанда күрделі өңдеу мен талдауды қажет етеді. Бұл сөзді тану, жазбаларды транскрипциялау және сөз тіркестері, сөздер мен дыбыстар сияқты сөйлеудің әртүрлі элементтерін анықтау қажеттілігіне байланысты болады.

Ауызша корпус – аудиомәтіндер мен бейнежазбалардың жинағы. Ауызша корпус не үшін қажет деген сауал туу мүмкін. Осы сұрауға жауап беріп көрейік:

1. Фонетика және фонология саласындағы зерттеулер: ауызша корпус сөйлеудің айтылуы мен дыбыстық құрылымын талдауға, фонетикалық және фонологиялық ерекшеліктерді анықтауға мүмкіндік береді. Сондай-ақ суперсегментті деңгейдегі тілдің көрінісін, ауызша тілдің сегментті және суперсегментті бірліктерін талдауға, интонацияның компоненттері мен бірліктерін дәл анықтауға мүмкіндік береді.

2. Грамматика саласындағы зерттеулер: ауызша корпус грамматикалық құрылымдардың қолданылуын зерттеуге, тіл грамматикасындағы ерекшеліктер мен өзгерістерді анықтауға, грамматикалық формаларды қолданудағы вариациялық мүмкіндіктерді анықтауға болады.

3. Лексика саласындағы зерттеулер: ауызша корпус сөздердің қолданылуын, олардың мағыналарын, colloquialism-дерін және тілдің басқа лексикалық ерекшеліктерін зерттеуге мүмкіндік береді.

4. Прагматика саласындағы зерттеулер: ауызша корпус коммуникативті стратегияларды, соның ішінде сөйлеу актілерін қолдануды, сыпайылықты, контекстік бейімделуді және ауызша қарым-қатынастың басқа аспектілерін талдауға мүмкіндік береді.

5. Әлеуметтік лингвистика саласындағы зерттеулер: ауызша корпус әртүрлі әлеуметтік топтардың тілдік мінез-құлқындағы айырмашылықтарды зерттеуге, әртүрлі контексттер мен әлеуметтік жағдайларда тілдің вариацияларын анықтауға мүмкіндік береді.

6. Оқу материалдары мен оқыту әдістемелерін әзірлеу: ауызша корпусты студенттерге нақты мысалдар арқылы ауызша қарым-қатынас пен сөйлеуді

түсіну дағдыларын дамытуға көмектесетін тапсырмалар мен жаттығулар жасау үшін пайдалануға болады.

Қорыта айтқанда, ауызша корпус – бұл тілді қолданудың нақты жағдайларында тіл мен коммуникацияның әртүрлі аспектілерін зерттеуге мүмкіндік беретін құнды зерттеу құралы. Ауызша корпус арқылы тілдің фонетика-фонологиясын, лексикасын, диалектілік ерекшеліктерді, грамматикалық құрылымдардың, әртүрлі жағдаятқа сәйкес сөйлеу дискурсын зерттеуге болады. Сондай-ақ тілдік құбылыстарды лингвистикалық немесе социолингвистикалық тұрғыдан зерттеуге құнды дереккөз болады.

Ауызша корпус – лингвистикалық немесе социолингвистикалық зерттеу үшін жиналған ауызша сөйлеудің аудио немесе бейне жазбаларының жинағы. Ауызша корпус зерттеушілерге тілдің әртүрлі контексттерде нақты қолданылуы туралы құнды деректер береді және оның құрылымын, функциялары мен эволюциясын жақсырақ түсінуге көмектеседі. Ол фонетика, фонология, лексика, грамматика, дискурс және т. б. сияқты әртүрлі салаларды зерттеуге арналған құнды деректер көзін ұсына алады. Ауызша корпус әртүрлі әлеуметтік және мәдени топтардағы айырмашылықтарды талдау үшін, уақыт өте келе тілдегі өзгерістерді зерттеу үшін, оқыту әдістемелерін әзірлеу және дағдыларды бағалау үшін пайдалануға болады.

2.5 Қазақ тілінің ауызша ішкорпусы: мәтіндер корпусындағы фонетикалық, просодикалық белгіленімдер әзірлемесі

Корпустық лингвистика компьютерлік технология арқылы тілді ақпараттандыру деген ұғымды береді. Қазіргі әлем тілдерінде технологияның дамуымен байланысты тілді компьютер бағдарламасы арқылы автоматтандыру мәселесі кең ауқымды дамып келеді. Жалпы «корпус» мәселесін ғылыми тұрғыдан зерттеген ғалымдар бар. Олар: Т. McEnery, Wilson [41], J.Svartvik [42], Ch. F. Meyer [43], К.П.Чилингарян [44], А.Жұбанов [12], Э.Сүлейменова, Г.Мәдиева, А.Фазылжан, А.Жаңабекова.

«Корпус» ұғымы туралы ғалымдардың ғылыми зерттеулеріне сүйенер болсақ, мысалы В.П.Захаровтың «Корпусная лингвистика» атты еңбегінде: «Под названием лингвистический, или языковой, корпус текстов понимается большой, представленный в электронном виде, унифицированный, структурированный, размечанный, филологический компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач» деп «корпус» ұғымын анықтап береді [45]. Осы секілді Э.Финеган, В.В.Рыков сынды зерттеушілердің еңбегінде де корпус туралы анықтама берілді.

Электрондық корпусты жасау және оны дамыту қазіргі тіл білімінің заманауи ең озық бағыттарының біріне жатады. Корпус дегеніміз арнайы бағдарлама арқылы компьютерде жинақталған деректер базасы ғана емес, сондай-ақ лингвистикалық тұрғыдан білімдер жүйесі болып табылады. Корпусты негіздеуде оның өзіндік әдістері бар. Корпустық лингвистика әдістері ішінен келесі бөліктерді жеке қарастыруға болады: 1) филологиялық әдістер; 2) теориялық-лингвистикалық әдістер; 3) математикалық (статистикалық)

әдістер; 4) ақпараттық технологиялық әдістер. Бірінші топ әдісіне жанрға, мәтін авторларына, жанр мәтіндерінің үлесіне және т.б. мәселелерге қатысты әдістер жатады. Екінші топта – мәтін таңдаудың және корпустағы лингвистикалық ақпараттың берілуінің жалпы және жеке лингвистикалық мәселелерімен байланысты әдіс-тәсілдер. Үшінші топты таңдама мәтіндердің санына, оның репрезентативтілігін (көлемділігін) қамтамасыз ету аспектілеріне қатысты математикалық статистиканың әдістері құрайды. Төртінші топқа корпус деректерінің компьютерлік көрінісі мен оны өңдеу әрекеттерін қамтамасыз ететін ақпараттандыру әдістері жатады. Корпустағы бірнеше түрі бар: жазбаша және ауызша (дыбыстық), параллель (бірнеше тілдің мәтіндері), стиль бойынша (публицистикалық, көркем әдебиет, іскери, ғылыми, ауызша), хронологиялық (синхронды, диахронды) т.б. Ұлттық корпус бірнеше ішкорпуста тұрады. Соның бірі ретінде ауызша ішкорпусты айтуға болады.

Қазіргі тіл білімінде ауызша корпус мәселесімен айналысып, оның тіл біліміндегі рөлін ғалымдар анықтап берді. Алайда қайсібір ауызша корпусты алып қарасаңыз да оның жасалуы мен орындалу деңгейі әртүрлі. Мысалы, орыс тіл білімінің ұлттық корпусында ауызша, акцентологиялық және мультимедиялық корпус түрлері бар. Орыс тілінің корпус базасындағы жазбаларда кіші, орта және жасы ұлғайған қарт кісілердің ауызша сөзі беріледі. Орыс тілінің ауызша және акцентологиялық ішкорпусында мәтінге сөз екпіні қойылып, орфографиялық норма бойынша жазу негізге алынған. Орыс тілінің ауызша корпус базасында 13 млн., акцентологиялық ішкорпуста – 133 млн., ал мультимедиялық ішкорпуста – 5,7 млн. сөз жинақталған. Ал британдық корпус базасының 10 пайызы ауызша ішкорпусқа негізделген. Ауызша ішкорпус екі бөлімнен тұрады. Бірінші демографиялық бөлімінде әртүрлі әлеуметтік топтағы адамдардың сөйлеу тілі мәтіндерінің транскрипциясы берілсе, ал екіншісінде іскер адамдардың сөзі немесе үкімет адамдарының сөздерінен бастап, радиошоулар мен телефон арқылы сөйлеген сөздер транскрипциясы беріледі. Барлық 10 миллиондық ауызша тілдің субкорпусы (шағын корпусы) шамамен бірдей екі бөлікке бөлінген: 1) «спонтанды» (дайындықсыз), табиғи диалогтар транскрипциясын сақтайтын демографиялық бөлік; 2) маңызды рөл атқаратын контекстік-басқарылатын бөлік. Бұл бөлікте іс-шаралардан жазылып алынған ауызша мәтіндер берілген. Ал чех ұлттық корпусының ауызша ішкорпусы базасы 4 млн. сөзқолданыстан тұрады. Неміс сөйлеу тілі мәтіндерінің корпусты жасау Германияда ертерек басталды. Сөйлеу тілі корпусты сақтаудың негізгі формасы магниттік таспаға жазып алу болды. Алайда кейінірек аудиожазба фонетикалық транскрипциямен берілді. Фонетикалық транскрипция пайдаланушыға түсініксіз болғандықтан, қиындық келтірді. Фонетикалық транскрипцияның қысқартылған нұсқасында интонацияны, айтылымның аймақтық ерекшеліктерін, сөйлеудің айқындылық дәрежесін және т.б. бейнелейтін құралдар қолданылды. Яғни әлемдегі тілдердің ауызша ішкорпусының зерттелуі мен қазіргі көрінісін қарастыратын болсақ, бірі орфографиялық норманы басшылыққа алса, ал кейбірі фонетикалық транскрипцияны негізге алды.

Сондай-ақ кейбір тілдерде жазба мәтіндермен қатар ауызша формадағы мәтіндерді де корпуста енгізеді. Корпуста енгізілген жазба мәтіндер мен ауыз-

ша мәтіндер сөйлеудің осы екі формасының арасындағы айырым белгілерді, өздеріне тән ерекшеліктерін айқындауға мүмкіндік береді. Жазба мәтіндер дәстүрлі ұстанымдарға, тілдік нормаға сәйкес жазылса, ауызша мәтін тілдің динамикалық сипатын танытады. Яғни ауызша мәтіндер – табиғи тірі тілдің бірден-бір көрінісі. Осы орайда қазақ тілінің ауызша ішкорпус базасына ауызша формадағы мәтіндер енгізілді. Ауызша формадағы мәтіндерге радиодан берілетін мәтіндер, сұхбаттар, дөңгелек үстел материалдары, қоғам қайраткерлерімен әңгімелесу сияқты ресми мәтіндер енді. Алынған мәтіндерге просодикалық белгіленімдер қойылды. Қажетті сөзді іздегенде ауызша корпус конкордансынан шығатын әрбір мысалдың аудиожазбасы/видеожазбасы қасында тұрады. Просодикалық белгіленімде мәтін синтагмаға, фразаға бөлінеді, паузасы анықталады және орфоэпиясы жасалады. Просодикалық деңгейдегі белгіленім үшін автоматтанған жүйе құру мәселесі әзірше өте күрделі болғандықтан, негізгі белгі-код қою жұмысы «қолмен» атқарылады.

Қай халықтың болса да, оның сөйлеу тілінің өзіне тән интонациясы бар. Интонация өзінің ішкі сипатымен ерекшеленеді. Бұндай ерекшеліктер мынадай просодикалық тәсілдерден тұрады: дауыс әуені, қарқыны, ұзақтылығы, кідірісі, үдемелігі, ырғағы, реңкі. Просодикалық тәсілдер сөзді дауыстың құлпыруына, өзгеруіне қарай түрлендіріп, мәнерлеп, көріктендіріп отырады, сөйлеушінің сөзіне ажар береді және айтылған ойды дамытады. Осы арқылы мәтіндегі сөйлемдердің мағынасы айқындалып анықтала түседі. Просодикалық тәсілдер мәтіндегі сөйлемдерді бөлшектеп, мүшеленген сегменттердің бір-біріне қатысын ашып, анықтайды және айтқан ойды түсіндіреді. Барлық тілдердегі интонацияның ең басты да универсалды просодикалық тәсілі – әуен. Себебі қай тілдің болса да басқа тілдерден айырмашылығын білдіруде сол тілдің өзіне ғана тән әуені үлкен рөл атқарады. Жалпы әуеннің сөйлемдегі атқаратын қызметі әралуан. Біріншіден, сөйлеу тілінің ағымын бөлшектерге мүшелеп жіктейтін де, олардың басын қосатын да әуен мен кідіріс. Екіншіден, сөйлемдердің коммуникативтік түрлерін анықтауда, яғни олардың айтылу мақсатына қарай (хабарлы, сұраулы, бұйрықты, лепті) бөлінуінде үлкен қызмет атқарады. Үшіншіден, грамматикалық және лексикалық тәсілдермен қосылып сөйлемдердің жалпы мағынасын беруге, оның ерекшеліктерін анықтауға қатысады. Төртіншіден, оның сөз тізбектерінің синтагмалық акценттелуімен қосылып, сөйлемнің логикалық және эмоциялық мазмұнын жеткізуде мәні зор. Дауыс әуені әр адамның сөйлеу мәнеріне қарай өзгеріп, сөйлеу ерекшеліктеріне қарай құбылып, стилистикалық бояумен сипатталуы мүмкін. Алайда әуеннің нормаға сәйкес негізгі формасы сақталады, себебі онсыз сөйлемнің негізгі мазмұны өзгеріп, түсініксіз болып, бір ойды жеткізе алмауы мүмкін.

Ауызша ішкорпустың просодикалық белгіленімін әзірлеуде мәтін синтагмаларға, ритмикалық топтарға бөлінді. Сонымен қатар просодикалық тәсілдердің ішінде әуеннің контуры (көтеріңкі, бәсең, көтеріңкі-бәсең, бәсең-көтеріңкі, бірқалыпты), кідірістің түрлері (эмоциялық, хезитациялық), екпіннің дискурстағы көрінісі (ой екпіні, эмоциялық екпін) басшылыққа алынды. Мәтінге просодикалық белгіленімдерді қоюда кідірістің орны айрықша. Себебі адам сөйлегенде барлық сөзді бір-біріне ұластырып айта бермейді, кейде

уақыт жағынан аздап тоқтайды, кейде ұзақ тоқтап барып сөзін жалғайды, яғни кідіріс жасайды. Дауыс кідірісінің негізгі міндеті мәтінді бөліп мүшелеу және сол бөлшектенген мүшелердің арақатынасын айқындау, мағынаны түсіндіріп, жазылған мәтінді ажыратып оқу, түсіну үшін белгілі дәнекерлік қызмет атқару. Кідірістің лингвистикалық функциясы дегеніміз – оның бір бүтін мәтінді фразаларға, синтагма, ырғақты топ, сөздерге бөлуі ғана емес, сонымен қатар интонацияның басқа компоненттерімен қосылып сөйлемдердің түрлі мазмұндық, эмоциялық т.б. қарым-қатынастарын анықтауға қатысуы [46].

Сонымен қатар просодикалық тәсілдердің біріне жататын үдемелікті де ауызша ішкорпуста анықтауға болады. Үдемелік басқа компоненттермен бірігіп барып бірашама мазмұнға ие болады. Үдемелік көбінесе дыбыстардың негізгі тон жиілігімен бірігіп, сөйлемдегі сөздердің акценттелуіне қатысады. Сөйлемдегі сөз тізбектерінің бір үзегін (сегментін) акценттеу керек болса, сол жердегі үдемелікті күшейтеді. Бірақ үдемеліктің күшейтілгені құлаққа көбінесе дауыстың қатты шыққанындай болып естіледі. Дауыстың қатты шығуы тек үдемелікпен ғана емес, негізгі тонының көтерілуімен де байланысты. Сөйлемдегі үдемелігі бірдей екі сегменттің біреуінің құлаққа қаттырақ естілуі, оның негізгі тонының жоғары екенін көрсетеді. Дауысты дыбыстардың үдемелігінің абсолюттік мәніне олардың артикуляциялық ерекшеліктері де әсер етеді. Қысаң дауысты дыбыстармен салыстырғанда ашық дауысты дыбыстардың үдемелігі мол екені экспериментті әдіспен дәлелденді.

Жоғарыда аталған просодикалық белгіленімдер ауызша ішкорпуста қолмен жасалады. Себебі компьютер бағдарламасына салатын болсақ, аудио-мәтіндерге қайта түзету жүргізу керек. Айталық, орыс тілінің ұлттық корпусында да просодикалық белгіленім негізінен қолмен жасалады. Оны авторлар былай түсіндіреді: «Просодическая разметка описывают ударение и интонацию. Такой разметке сопутствует дискурсная разметка, служащая для отметки повторов, оговорок и так далее. Разметку текстов делают при помощи программных средств, что позволяет снизить трудозатраты. Для анафорической и просодической разметок создание подобных программных средств является сложной задачей, поэтому большую часть работы проводят вручную. Применяемые программные средства требуют постредактирования, которое происходит вручную (морфологическая омонимия и синтаксическая неоднозначность), так как программы представляют некоторое количество вариаций решения, а нужное выбирает сам исследователь. В перспективе находится полный автоматизированный процесс разметки» [47].

Қазақ тілінің ауызша ішкорпусына әзірленіп жасалатын мәтіндерге қойылатын просодикалық белгіленімдердің өзіндік ерекшелігі бар. Оның бірі мәтіндердегі пунктуациялық белгілердің орнына қысқа кірідірістің (/), синтагманың (//) және фразаның (///) шартты белгілері қойылады. Осы арқылы корпусты пайдаланушы ауызша мәтіннің мазмұнын бірден түсінеді. Просодикалық белгіленім негізінен арнайы бағдарлама арқылы емес, қолмен жасалатындықтан, ондағы мәтіндер ауқымы да аз қамтылады. Просодикалық белгіленімді жетілдіру барысында сөйлеуші сөзінің орфоэпиялық нормасын көрсету үшін оның ауызша мәтіні жартылай автоматқа салынады. Автомат

ауызша мәтіндегі кейбір дыбыстарды алгоритм бойынша орфоэпиясын жасап шығады. Алгоритм қазақ тілінің орфоэпиялық нормасы бойынша жасалды.

Ауызша корпустың базасына просодикалық белгіленімді енгізуде ең алдымен ғылыми ізденістермен (просодикалық белгілерді қоюдың ғылыми негіздерін) тетіктерін тауып, зерттеп алғанымыз жөн. Орыс тілі корпусының просодикалық белгіленімінде тек сөйлемдер синтагмаға бөлініп, екпін ғана қойылған. Ал қазақ тілінің ауызша ішкорпусының ерекшелігі – әрбір сөйлемнің фонетикалық транскрипциясын және просодикалық белгіленімін қойып, сөйлемді жеке синтагмаларға бөліп, интонемаларын (көтеріңкі, бәсең, бірқалыпты) беруде. Сөйлеуде (мәтінде) көрініс табатын ырғақтық топтар мен синтагмалар фонетикалық және семантикалық-интонациялық тұрғыдан мағына ажыратуға негіз болады [48]. Яғни әрбір сөйлем іштей синтагмаларға бөлініп, сөйлеудің (мәтіннің) семантикалық-интонациялық ерекшелігі көрсетіледі. Синтагма дегеніміз барлық тілдерге тән әмбебап, айтылымды, мәтінді сегменттеу кезінде пайда болатын семантикалық бірлік болып табылады. Мәтінді синтагмаларға бөлу кезінде сөйлемнің мағынасы бұзылмауы үшін синтагманың шекарасын дұрыс қою маңызды. Синтагмалардың шекарасын қою семантикалық мазмұнның берілуімен қатар интонациялық сипаттардың берілуіне де әсер етеді. Мысалы: Балалар жазда | демалысқа шығады. – деген сөйлем екі синтагмадан тұрып, алдыңғысы белгілі ақпаратты көрсететін тема болса, ал соңғысы жаңа ақпаратты білдіруші рема болып табылады. Осы екеуі сөйлемді байланыстырып, мағыналы сөйлем құруға көмектеседі. Алдыңғы синтагма құрайтын темаға жоғары көтеріліп, мазмұнның аяқсыз екенін көрсететін әуен көрсеткіші тән болса, ремаға ойды аяқтап, төмен түсетін әуен тән.

Ауызша ішкорпус базасындағы ауызша сөздің просодикалық белгіленімін қою үшін, яғни әрбір сөйлемнің интонациясын анықтау үшін академик З.М.Базарбаеваның, А.Фазылжанның зерттеулері басшылыққа алынды. Академик З.М. Базарбаева қазақ тіл білімінде интонацияның негізгі бес функциясы бар екенін анықтады. Олар: сөйлемнің жалпы коммуникативтік түрлерін анықтау (сөйлемнің хабарлы, сұраулы, бұйрықты түрлерін ажыратады); жеке коммуникативтік типтерді ажырату (сөйлемді тема мен ремаға бөлу); мазмұнның өзгешеліктерін ажырату (мазмұнынан қарай, санамалау интонациясы, қарсылықты, біріктіру, жалғастыру, қыстырма сөз, сөйлем интонациясы); сөйлемді, синтагмаларды, мәтінді бір-бірінен ажырату (сөйлем түрлерін дұрыс мәнерлеп, олардың жіктерін ажырату, сөздерді анық айту.); эмоцияны айқындау (реңі ғана емес, мағынасының өзгеруі). Ғалым: Мен сізді жақсы білемін деген сөзді интонация арқылы әртүрлі (кекету, риза болғаннан айту) мағынада беруге болатынын айтады [49]. Просодикалық белгіленімде интонациялық компоненттердің бірі әуен басшылыққа алынды. Сондай-ақ ауызекі сөйлеуде екпіннің де қызметі ерекше. Екпін қызметі әр тілде әртүрлі көрініс табады. Соның ішінде орыс тіліндегі екпіннің жылжымалылығы (басында да, соңында да келе беруі) оны флективті тілдер қатарына қосса, қазақ тілінде екпіннің көбінесе сөздің соңғы буынына түсуі қазақ тілінің жалғамалы (агглютинативті) тілдер қатарында екендігін көрсетеді. Яғни І. Кеңесбаев айтқандай [50], бұл екі тілдің типологиялық (қазақ тілі – агглютинативті, орыс тілі – флективті)

өзгешелігіне байланысты екенін танытып, ғалымдар жан-жақты дәлелдеген. Осы негізде Ә. Жүнісбектің пікірінше, қазақ тіліндегі сөздердің бәрі бірдей сөйлем ішінде үнемі дербес екпінге ие бола бермейді [51].

Просодикалық белгіленім жаңадан жасалып жатқанымен, ҚТҰК-да фонетика-фонологиялық белгіленім толық жұмыс жасайды. Ұяшықтағы кез келген сөзді бассаңыз, осы сөздегі әрбір фонемаға фонологиялық сипаттама беріледі. Сөз ағымында әртүрлі дыбыстар айтылады, олардың айтылу түрі көп, шегі жоқ, бірақ олар орайласып, бір дыбыс типін құрайды, оны біз фонема дейміз. З.М.Базарбаева фонемаға қатысты мынадай анықтама береді: «Тілдің ең шағын функционалды бірлігі, сөз мағынасы мен сөз тұлғасын ажырататын және оның құрамында болатын дыбыс типі». Фонемалар мағына айқындау жағынан жұп-жұбымен қосарланып, бір-біріне қарама-қарсы қойылады, сонда оның негізгі қасиеті, ерекшелігі анықталады. Бір сөзде фонемалардың саны аз да, көп те болуы мүмкін, бір сөз екінші бір сөзден өзгеше болуы үшін барлық фонеманың өзге болуы шарт емес. Бір сөз басқа сөздерден фонемасының саны жағынан, сапасы жағынан және сөздегі орын тәртібі жағынан өзгешеленеді [52]. Мысалы, <А> фонемасы сөздің барлық позициясында кездеседі. Барлық дауыссыз дыбыспен тіркеседі. [Ш] мен [й]-дің, [ж] мен [й]-дің және жіңішке буыннан кейін [й] мен іргелес жіңішке буынның алдында келгенде, <а> фонемасының дыбыстық репрезентанты [ә] түрінде болады, яғни палатализацияланады. Енді осы <а> фонемасының [ә] вариантымен келетін сөздерді тіземіз. Ж<а>й – ж[ә]й, ш<а>й – ш[ә]й, м<а>йсөк – м[ә]йсөк, бид<а>й – бид[ә]й, ауж<а>й – [ә]уж[ә]й, д<а>укес – д[ә]укес, д<а>укестік – д[ә]укестік, ж<а>йбарақат – ж[ә]йбарақат, ж<а>йбарақаттану – ж[ә]йбарақаттану, ж<а>йбасар – ж[ә]йбасар, ж<а>йғастыру – ж[ә]йғастыру, ж<а>йғасу – ж[ә]йғасу, ж<а>йғату – ж[ә]йғату, ж<а>йғызу – ж[ә]йғызу, ж<а>йдақ – ж[ә]йдақ, ж<а>йдарлы – ж[ә]йдарлы, ж<а>йдарман – ж[ә]йдарман, ж<а>йдары – ж[ә]йдары, ж<а>й-күй – ж[ә]й-күй, ж<а>йқалу – ж[ә]йқалу, ж<а>йлы – ж[ә]йлы, ж<а>йма – ж[ә]йма, ж<а>йма-шуақ – ж[ә]йма-шуақ, ж<а>йсаң – ж[ә]йсаң, ж<а>йша – ж[ә]йша, ж<а>йшылық – ж[ә]йшылық, ж<а>йсыз – ж[ә]йсыз, ж<а>йнау – ж[ә]йнау, ж<а>йт – ж[ә]йт. Сонымен қатар дауыссыз [к] дыбысынан кейін дауысты <а> фонемасының дыбыстық репрезентанты [ә] түрінде болады, яғни палатализацияланады: қ<а>дір – қ[ә]дір, қ<а>дірлеп-қастерлеу – қ[ә]дірлеп-қ[ә]стерлеу, қ<а>дірлеу – қ[ә]дірлеу, қ<а>дірлі – қ[ә]дірлі, қ<а>жет – қ[ә]жет, қ<а>жетсіз – қ[ә]жетсіз, қ<а>стерлеу – қ[ә]стерлеу, қ<а>стерлі – қ[ә]стерлі. ... [к]-ның әдеттегі жуан тембрінен басқа жіңішке тембрлі болып келуі түрленім деңгейінде үндесім деуге болады. Сондай-ақ қауесет дегендегі [ә] соңғы буындағы [е]-нің әсерінен жіңішке реңкпен айтылады. Бірақ бұл дауыстының естілімі кәдімгі [а]-мен әдеттегі [ә]-нің аралығы деуге болады (Орфоэпиялық сөздік). Дауыссыз [н] мен [с] аралығында <а> фонемасының дыбыстық репрезентанты [ә] түрінде болады: н<а>сихат – н[ә]сйхат, н<а>сихатшы – н[ә]сйхатшы, н<а>сихаттау – н[ә]сйхаттау.

Жіңішке кірме сөздердің II буынында <а> фонемасының дыбыстық репрезентанты [ә] түрінде болады: ділд<а> – ділд[ә], тілм<а> ш – тілм[ә]ш, тілм<а> ш – тілм[ә]штық, дінд<а> р – дінд[ә]р.

Жіңішке кірме сөздердің III буынында <a> фонемасының дыбыстық репрезентанты [ə] түрінде болады: дүб<a>р<a> – дүб[ə]р[ə].

Фонемалардың төменде берілген сипаттамасы ауызша ішкорпусты жасауға негіз болды.

Кесте 2

Қазақ тіліндегі фонемалардың сипаттамасы

№	фонема	Сипаттамасы
1	<a>	ашық, езулік, тіларты; жуан дауысты фонема.
2	<ə>	ашық, езулік, тілалды; жіңішке дауысты фонема.
3	<ы>	қысаң, езулік, тіларты; жуан дауысты фонема.
4	<і>	қысаң, езулік, тілалды; жіңішке дауысты фонема.
5	<ү>	қысаң, еріндік, тіларты; жуан дауысты фонема.
6	<у>	қысаң, еріндік, тілалды; жіңішке дауысты фонема.
7	<е>	жартылай ашық, езулік тілалды; жіңішке дауысты фонема.
8	<o>	ашық, еріндік, тіларты; жуан дауысты фонема.
9	<ө>	ашық, еріндік, тілалды; жіңішке дауысты фонема.
10	<п>	ерін-ерін нық тоғысыңқы тұтас қуысты; қатаң дауыссыз фонема.
11	<б>	ерін-ерін босаң тоғысыңқы көмей қуысты; ұяң дауыссыз фонема.
12	<м>	ерін-ерін бос тоғысыңқы ауыз қуысты мұрын жолды; үндідауыссыз фонема
13	<т>	тіл ұшы нық тоғысыңқы көмей қуысты; қатаң дауыссыз фонема.
14	<д>	тіл ұшы босаң тоғысыңқы көмей қуысты; ұяң дауыссыз фонема.
15	<н>	тіл ұшы, босаң тоғысыңқы көмей қуысты мұрын жолды; үнді дауыссыз фонема.
16	<қ>	тілшік нық тоғысыңқы көмей қуысты; қатаң дауыссыз фонема.
17	<к>	тіл ортасы нық тоғысыңқы ауыз қуысты; қатаң дауыссыз фонема.
18	<ғ>	тілшік босаң тоғысыңқы көмей қуысты; ұяң дауыссыз фонема.
19	<г>	тіл ортасы босаң тоғысыңқы ауыз қуысты; ұяң дауыссыз фонема.
20	<ң>	тілшік бос тоғысыңқы көмей қуысты мұрын жолды; үнді дауыссыз фонема.
21	<с>	тіл ұшы жинақы жуысыңқы көмей қуысты; қатаң дауыссыз фонема.
22	<з>	тіл ұшы босаң жуысыңқы көмей қуысты; ұяң дауыссыз фонема.
23	<р>	тіл ұшы бос жуысыңқы көмей қуысты ауыз жолды діріл; үнді дауыссыз фонема.
24	<ш>	тіл ұшы, нық жуысыңқы көмей қуысты; қатаң дауыссыз фонема.
25	<ж>	тіл ұшы жайылыңқы жуысыңқы көмей қуысты; ұяң дауыссыз фонема.
26	<л>	тіл ұшы бос тоғысыңқы көмей қуысты ауыз жолды; үнді дауыссыз фонема.
27	<й>	тілшік бос жуысыңқы көмей қуысты ауыз жолды; үнді дауыссыз фонема.
28	<у>	ерін-ерін бос жуысыңқы көмей қуысты ауыз жолды; үнді дауыссыз фонема.
29	<ф>	тіс-ерін нық жуысыңқы; қатаң дауыссыз бөгде фонема.

30	<ц>	тіл ұшы қызыл иек аффрикат жабысыңқы нық жиылыңқы; қатаң дауыссыз бөгде фонема.
31	<ч>	тіл ұшы қызыл иек аффрикат нық жабысыңқы-жайылыңқы; қатаң дауыссыз бөгде фонема.
32	<h>	тілшік жинақы жуысыңқы көмей қуысты; қатаң дауыссыз фонема.
33	<x>	тілшік көмей қуысты; қатаң дауыссыз фонема.
34	<в>	тіс-ерін босаң жуысыңқы ауыз қуысты; ұяң дауыссыз бөгде фонема.
35	/я/	[й+а]; [й+ә] дыбыстарының қосындысын білдіреді.
36	/ю/	[й+ү+у]; [й+ү+у] үш дыбыстың қосындысын білдіреді.
37	<э>	ашық, езулік, тіларты; жуан дауысты бөгде фонема.
38	/ш/	[ьй], [ій] дыбыс тіркесінен тұрады.
39	/ш/	Бөгде сөздерде ауыз қуысты жіңішке тіл алды; ызың қатаң дауыссыз фонема. Төл сөздерде қос шш дыбысын береді.

Осы үлгіде берілген фонемалардың сипаты ұлттық корпусның фонетика-фонологиялық белгіленім бөлігінде нақты көрсетілді. Кез келген сөзді теріп, оның үстінен басқанда фонетика-фонологиялық сипаты беріледі.

Жоғарыда айтылған тұжырымдарды жинақтай келе, ауызша ішкорпусқа енетін мәтіндердің бірі үлгі ретінде беріліп отыр. Зерттеуге Ә.Нұршайқовтың «Ақиқат пен аңыз» атты шығармасынан үзінді алынып, жоғарыда айтылған просодикалық белгіленімдер қойылды.

- 1-қадам сөйлеушінің сөзі орфографияға сәйкес берілді.

Сонымен, мың тоғыз жүз отыз үшінші жылдың аяғында, қақаған қыста, шинелімнің екі етегі делендеп, шлемімнің екі құлағы салпаңдап, Бурныйға тоқтаған пойыздан дік етіп жерге түстім. Пойызда келе жатқанда Тайганың алдан қашқан ақ түлкісінің құйрығындай жеп-жеңіл, жұп-жұмсақ сияқтанып, дала бетінде бұлаңдап тұрған боранның жерге түскенде екпіні ер жігітті алып соғардай қатты екен. Маңдайында қызыл жұлдызы жарқыраған шошақ шлемнің екі құлағын иегіме қамзау етіп байладым да, алдымен ауыл жаққа қарадым. Ауыл көрінбейді. Мынау сансыз ақ түлкінің құйрығындай бұлаңдаған ақ түтектердің ар жағында, төрт-бес шақырым жерде ғана тұр. «Жүр-жүрлеп!» жүректі сағыныш жетелейді. «Сабыр, сабыр!» деп санам тартпақтайды. Жас адамға жүрек би. Сана билейтін шар тартқан кез емес еді ғой ол. Сақылдаған аяз, соғып тұрған боранға қарамастан: «Қайдасың аяулым, ауылым? Қайдасың ата-ана, бауырым?!» – деп, *Мыңбұлақты бетке алып, жүрдім* де кеттім.

- 2-қадам сөйлеушінің өз сөзі просодикалық белгінім бойынша берілді.

>сонымен мың тоғужүз отыз үшінші → *жылдың айағында* | <↑ – қақаған қыста <↑ | – шинелімнің → | ек' етегі делендеп – <↑ | ш'лемімнің → | ек' құлағы → салпаңдап – <↑ | бурныйға тоқтаған пойыздан → > | *дік етіп* ↑ | – жерге түстім > ↓ || *пойызда келе жатқанда* → > | *тайганың* → | *алдан қашқан* → | *ақ түлкісінің құйрығындай* ↑ < | жеп-жеңіл > → | *жұп-жұмсақ сияқтанып* – <↑ | дала бетінде → > | *бұлаңдап тұрған боранның* ↑ < | жерге түскенде → | *екпіні* → < | *ер жігітті алып соғардай* ↑ < | *қатты* екен ↓ || *маңдайында* ↑ < | *қызыл жұлдызы жарқыраған*

→ | шошақ шілемнің → < | ек' құлағын → | ійегіме қамзау етіп ↑ < – | байладым да > ↑ | алдымен → < | ауұл жаққа қарадым > ↓ | ауұл Гөрүмбейді > ↓ | мынау сансыз → > | – ақ түлкүнің құйрығындай ↑ < | бұлаңдаған ↓ > | ақ түтөктердің < ↑ | әржәгьында – → | төрт-пес шақырым → | жерде гана тұр > ↓ || жүр-жүрлөп < ↑ – | жүрөкті < → | сағыныш жетелейт. > ↓ || сабыр | сабыр деп ↑ < – | санам тартпақтайт > ↓ | жас адамға → < – | жүрөк – бій > ↓ | сана бійлейтін → < | шар тартқан → – | кез емес → | еді гой ол > ↓ || сақылдаған айаз → > | соғып тұрған → > | бораңға қарамастан < ↑ «қайдасың айаулым – ↑ | ауулым < ↑ қайдасың < – ↑ | ат'ана → бауұрұм < ↑ – деп < → | мыңбұлақты бетк'алып – < → | жүрдүм де кеттім > ↓

• 3-қадам сөйлеушінің сөзі қазіргі орфоэпиялық нормаға салынып, просодикалық белгінімі берілді.

> сонұмен → | – мың тоғұжжүз отүз үшүнушү → жылдың айағында | < ↑ – қақаған зыста < ↑ | – шійнелімнің → > | "ек' тегі делеңдеп – < ↑ | шілемінің → > | "ек' құлағы → салпаңдап – < ↑ | бұурныйға тоқтаған пойыздан → > | дік "етіп | ↓ – жерге түстүм > ↓ || пойызда жележатқанда → > – | тайганың → | алдаң ғашқан > → | ақ түлкүсүнің құйрұғұндай ↑ < | жеб-жеңіл > → | жұб-жұмсақ сыйақтанып – < ↑ | дала бетінде → > | бұлаңдап тұрған боранның ↑ < | жерге түскөнде → – ↓ "екпіні → < | "ер жігітті алып соғардай < ↑ – | қатт'екен ↓ || маңдайында ↑ < – | қызыл жүлдүзы жарқыраған → | шошақ шілемнің → < | "ек' құлағын → | ійегіме қамзау етіп ↑ < – | байладым да > ↑ | алдымен → < | ауұл жаққа ғарадым > ↓ | ауұл гөрүмбейді > ↓ | мынау сансыз → > | – ақ түлкүнің құйрұғындай ↑ < | бұлаңдаған ↓ > | ақ түтөктердің < ↑ | аржәгьында – → | төрт-пес шақырым → | жердегана тұр > ↓ || жүр-жүрлөп < ↑ – | жүрөкті < → | сағыныш жетелейді > ↓ || сабыр | сабыр деп ↑ < – | санам тартпақтайды > ↓ | жас адамға → < – | жүрөк – бій > ↓ | сана бійлейтін → < | шар тартқан → – | кез "емес → | "едігой ол > ↓ || сақылдаған айаз → > | соғып тұрған → > | бораңға қарамастан < ↑ қайдасың айаулым – ↑ ↓ ауулым < ↑ қайдасың < – ↑ | ат'ана → бауұрұм < ↑ – деп < → | мыңбұлақты бетк'алып – < → | жүрдүм де ғеттім > ↓

Зерттеуде корпус базасына интонация белгілерін беру үшін арнайы белгілер қойылды.

Әрбір интонеманың сөйлемнің басында көтеріңкі, ал аяғына қарай бәсең интономамен айтылатындығы арнайы белгілермен берілді. Сондай-ақ мәтіннің интонациясын беру үшін арнайы шартты белгілер қойылды. Яғни қазақ тілінің интонациялық ерекшелігін, яғни әуенін көрсететін 8 интонома мен дауыс үдемелігі, кідіріс, қарқын көрсеткіштері арнайы белгілермен интонациялық транскрипциясы берілді [53].

Корпус базасына енетін мәтіндердің просодикалық ерекшелігін көрсету үшін төмендегідей шартты белгілер қолданылды.

Интонациялық транскрипцияның белгілері: Сөйлеу кезіндегі синтагма аралары тік сызықпен (|), фраза арасы қосарланған (||) тік сызықпен белгіленді.

1. Фразадағы кідірісті көлбеу сызықпен (–) белгілеу.

2. Сегменттердің негізгі тон сызығын графикалық түрде белгілеу. Негізгі тон жиілігінің қозғалысы үш түрлі бағытта көрінеді:

- Көтеріңкі ↑
- Бәсең ↓

- Бірсыдырғы →

1. Сегменттердің айтылу қарқынын анықтау.

Қарқынның белгіленуі:

- Баяу
- Орташа
- Жылдам

4. Дауыс үдемелігін айқындау.

- Күшею <
- Азаю >

Зерттеу барысында мәтіннің интонациялық транскрипциясы аудиторлық талдау негізінде жүргізілді. Сондай-ақ әуені, дауыс үдемелігі, қарқыны, кідірісі анықталып, интонациялық талдау жасалды. Әрбір интонациялық компоненттерінің сөйлемнің мағынасын ашып беруде және ойды нақты жекізуде рөлі ерекше екендігін дәлелдеу үшін мысалдар келтіріледі. Мысалы:

<сонұмен | мың тоғұжжүз отұз үшүншү жылдың айағында | қақаған ғыста | шійнелімнің | ек' етегі делеңдеп | ш'лемімнің | ек' ғұлағы салпаңдап < | *бұрныйға тоқтаған пойыздан* | *дік етіп* | – жерге түстүм > – деген сөйлем он синтагмадан тұрады.

Әуен мен үдемелік. Бірінші, екінші синтагмада бірқалыпты интонемамен басталған сөйлем үшінші және төртінші синтагмада көтеріңкі яғни тиянақсыз интонемамен айтылып, бесінші, алтыншы синтагмада бірқалыптыдан қайта көтеріңкі айтылып, сөз аяғында бәсеңдеу байқалады. Яғни тұтас бір ойдың аяқталғандығын білдіреді. Әрбір интонема ойды дәл жеткізуде өзіндік рөлге ие.

Қарқыны. Сөз басы баяу басталып, үшінші мен төртінші синтагмада күшейгендігі, яғни ой екпінінің түсуінен қарқынның күшейгендігі, ал сөйлемнің соңында баяулағандық байқалады. Жалпы қарқын сөздің маңызды, маңызсыз тұстарын ерекшелеп көрсету, ойды нақтылау үшін қажет.

Кідіріс. Сөйлемде бес кідіріс жасалған. Әрбір кідіріс белгілі бір ойдың тиянақталуын, жеке-жеке синтагма арасындағы ойдың мағынасын ажыратуда да өзіндік рөлге ие болады. Кідіріс бір жағынан физиологиялық тұрғыдан жасалса, ал екінші жағынан логикалық тұрғыдан ойдың жинақылығын көрсетеді.

<маңдайында –| *қызыл жұлдызы жарқыраған* | *шошақ шілемнің* | *ек' Ғұлағын* | *ійегіме қамзау етіп* –| *байладым да* > | алдымен| ауұл жаққа ғарадым > – деген сөйлемде сегіз синтагма бар.

Әуен мен үдемелік. Сөйлемнің басында бірінші синтагма көтеріңкі интонемамен басталып, екінші мен төртінші синтагма аралығы бірқалыпты айтылған. Бесінші мен алтыншы синтагма тиянақсыз, яғни көтеріңкі әуенмен айтылғандығы байқалады. Сонан соң бірқалыпты интонемамен жалғасып, соңында бәсеңдей түскен. Бұл сөйлемде интонемалар арқылы айтылар ой нақты жеткізілген.

Қарқыны. Сөйлемнің басында қарқыны баяу басталып, ортасына қарай қарқыны күшейе түседі де аяғына қарай баяулай түскен. Эксперименттік зерттеулерден байқалатындығы, қарқын көрсеткішінің баяу айтылуы айтылар ойдың маңызды тұстарын көрсетсе, ал қарқынның жылдамдауы маңызды емес екенін байқатады.

Кідіріс. Бұл сөйлемде жеті кідіріс жасалған. Бұл бір жағынан физиологиялық кідіріс болса, ал екінші жағынан сөйлемнің мағынасын ашу мақсатында және ойдың толық мәнін ашу мақсатында екендігі байқалып отыр.

Интонациялық жағынан интонемалардың ерекшелігін көрсету мақсатында зерттеу жүргізіліп қана қоймай, просодикалық белгіленім жасауда сөздер мен сөз тіркестерінің орфоэпиясы беріліп, сөйлем синтагмаларға бөлініп, интонемалары қойылды. Мүмкін болған жағдайда қарқыны мен кідірісі де жасалды.

Просодикалық белгіленімді бірден автоматтандыру мүмкін емес. Ауызша корпуста мәтін автоматтандырылу үшін алгоритм беріледі, алайда ол ауызша мәтінді толық қамти алмағандықтан, қолмен жүзеге асырылады. Ал кейіннен базаның ауқымы кеңейіп, алгоритімі жетілдірілген сайын мәтін толық автоматтандырылады.

Қорыта келгенде, қазақ тілінің алғашқы ұлттық корпусында неғұрлым жеңіл әрі қарапайым белгіленімдер көрсетіліп, көлемі мейлінше аз (10 млн.) мәтін қамтылды. Осыған орай қазіргі таңда корпустың ауқымы, ақпараттар кеңдігі ұлғайып, белгіленім сапасы мен барлық параметрі мен подкорпус ресурстары толықтырылуда.

Просодикалық белгіленім негізінен ауызша мәтіндерден тұрады. Классик жазушылардың, қазіргі ақын-жазушылардың, қоғам қайраткерлерінің аудиожазбасы басшылыққа алынды. Ауызша мәтіндерде сөйлеушінің орфоэпиясы, просодикалық ерекшелігі, интонациялық жағынан безендірілуі нақты көрсетілді. Ұлттық корпусты пайдаланушы қазақ тілінің нормаға түскен орфоэпиясы мен сөйлеушінің орфоэпиясы арасында айырмашылықты ажырату үшін, сонымен қатар сөздің дұрыс дыбысталуының қандай деңгейде екенін айқындап алу үшін просодикалық белгіленім берілді. Пайдаланушы нормаға түскен орфоэпия мен сөйлеушінің орфоэпиясын қалай салыстырады деген сауалға келетін болсақ, просодикалық белгіленімі қойылған ауызша мәтіндегі сөздің үстінен басқанда оның нормаға түскен орфоэпиясы, интонациялық ерекшелігі көрсетіледі және оның аудиосы қатар беріледі.

Ұлттық корпустың бір міндеті просодикалық белгіленімді көрсету дейтін болсақ, оған қандай тапсырмалар кіреді. Осы жағынан келгенде ең бірінші мәтіннің орфоэпиясын автоматтандыру үшін дауысты және дауыссыз дыбыстардың алгоритмі жасалды. Алгоритмге орфоэпия бойынша бірнеше жылдан бері зерттеліп жүрген ғылыми еңбектердің нәтижесі, дауысты және дауыссыз дыбыстардың моделі басшылыққа алынды.

Зерттеуге фонетиканың сегментті жағымен қатар суперсегментті фонетика, яғни просодикалық жағы да қарастырылып, сипаттамалы, аудиторлық және формалды-сипаттамалық әдіс қолданылады. Формалды-сипаттамалық әдіс бұл интонацияның графикалық ұғымымен тікелей байланысты болады. Интонациялық транскрипциялардың әртүрлі болуы мүмкіндігінше түпнұсқаға жақын сөйленістің моделін қалыптастыруға мүмкіндік береді.

Просодикалық тәсілдер сөзді дауыстың құлпыруына, өзгеруіне қарай түрлендіріп, мәнерлеп, көріктендіріп отырады, сөйлеушінің сөзіне ажар береді және айтылған ойды жандандырып, мәтіндегі сөйлемдердің мағыналары, қарым-қатынастары айқындалып, анықтала түседі. Просодикалық тәсілдер

мәтіндегі сөйлемдерді бөлшектеп, мүшеленген сегменттердің бір-біріне қатысын ашып, анықтайды және айтқан ойды түсіндіреді.

Қазақ тілінің ауызша ішкорпусы жылдар өткен сайын мәтіндермен толықтырылып, жетілдіріліп отырады. Оны автоматтандыру үшін Praat бағдарламасы мен алгоритм қатар жұмыс жасап, базаның аясын кеңейте түседі. Алгоритм осы кезге дейін қазақ тіл білімінің сегментті және суперсегментті фонетика саласындағы зерттеулердің негізінде жасалып, мәтінді автоматтандыруға негіз болады.

2.6 Ауызша ішкорпус базасындағы мәтіннің фонетика-фонологиялық және интонациялық транскрипциясы

Қазіргі қазақ тіл білімінде фонетикалық транскрипция мен фонологиялық транскрипцияның ара-жігі ажыратылып, ерекшелігі анықталған болатын. Ғалымдар дыбыс пен таңбаның бір-бірімен үнемі сәйкес келе бермейтіндігін, яғни сөздің жазылуы мен айтылуына қатысты пікір білдіріп, лингвистикалық тұрғыдан арнайы транскрипция қолданылатынын айтады [54]. Фонетикалық транскрипция ауызша айтылған сөздің дыбысталуын неғұрлым дәл беруді көздесе, ал фонематикалық транскрипцияда фонеманың негізгі реңкі таңбалады.

Жалпы мәтінді транскрипциялауда дыбыстық ерекшеліктермен қатар просодикалық көрінісін де берген жөн. Яғни ол интонациялық транскрипция арқылы жүзеге асады. Жалпы ғалымдар интонация мен просодика ұғымын бірге қарастырады. Алайда іштей айырмашылығы да бар. Мысалы, просодиканың ұғымы интонацияға қарағанда кеңірек. Яғни просодикада буыннан бастап, сөз, ритмикалық топ, синтагма, сөйлесім деңгейін қамтыса, ал интонацияда синтагма мен сөйлесімді (высказывание) ғана қамтиды. Интонациялық транскрипция туралы айтпас бұрын ең алдымен интонацияның зерттелу тарихына тоқталған жөн.

Интонация – күрделі құбылыс. Зерттеушілер зерттеу жасау барысында интонацияның дыбыстан бұрын пайда болғандығын анықтаған. Оған дәлел ретінде хайуанаттардың қарым-қатынас жасауындағы дауыс интонациясының құбылуын байқауға болады. Сондай-ақ, жаңадан туған нәрестенің анасының дауысын, интонациясын сезе алуы, яғни анасының ұрысқаны мен еркелеткенін білуі дәлел бола алады. Осы орайда академик З.М. Базарбаева интонацияның адам сезімімен байланыстылығын, дауыс ырғағына байланысты қалыптасатын күрделі құбылыс екенін айтады [49]. Осымен байланысты А.М. Пешковскийдің де зерттеулерінде сөзден гөрі интонацияның атқаратын қызметінің ерекше екені көрсетілген [55].

Ғалымдар интонацияның өте ерте кезден басталғанын айтып, оны сахна өнерімен және шешендердің сөздерімен байланысты қарайды. Интонация туралы алғашқы дерек Греция мен Римде пайда болып, сөйлеу актісінде әуен, ырғақ, қарқынның болатыны көрсетілген.

Интонацияның ең алғаш теориялық жағынан зерттелуі XIX ғасырдан бастау алады. Г. Суит және Д. Джоунз интонацияның белгілі бір функция

атқаратындығын ойлап, ең негізгі функциясы ойдың тиянақталуы мен тиянақталмауы дейді [56]. XX ғасырдың 70 жылдарына дейін интонацияны зерттеуде сипаттамалы әдіс қолданылды. Сипаттамалы әдіс интонация компоненттерін анықтауға, әсіресе әуеннің контурын анықтауға негіз болды. Кейін келе тек сыртқы сипатын қарастырып қана қоймай, оның табиғи заңдылығын анықтау мақсатында инструменталды әдіс арқылы зерттеу қолға алынды. Ең алғаш рет ғалымдар интонациялық транскрипцияны мысалдар арқылы көрсеткенде тек әуенді ғана бере алған.

Алғаш рет интонацияны қағаз бетіне салған ағылшын ғалымы Д. Стил болатын. Осы орайда ғалымның «Prosodia Rationalis» атты еңбегі жарық көрді. Ғалым еңбегінде тонның жоғарылығын, дыбыстың ұзақтылығын, буынның екпінділігі мен екпінсіздігін, дыбыстың акустикалық сипаты мен кідірісті көрсету үшін ноталық символдарды ұсынады. Ғалым ноталық символды өлең жолдарында қолдана отырып, ондағы екпінді, кідірісті және дауыс үдемелігін белгілеп берді. Осы ізбен орыс, испан, неміс ғалымдары да ноталық символды қолданды. Алайда ғалымдар кейінгі ғылыми-зерттеу барысында ноталық транскрипцияның интонацияны оқыту әдістемесінде де, ғылыми-зерттеу талдауларда да тиімсіз екенін көрсетті. Ноталық транскрипция тек әуеннің сан қырлы реңктерін анықтау үшін өте тиімді әдіс болды. Ғалымдар ноталық транскрипцияны беру үшін ноталық білімнің де қажет екенін алға тартты. Кейін келе акустикалық құралдар пайда болып, дамуының арқасында инструменталды зерттеулер жүргізуге мүмкіндік туды да арнайы фонетикалық зертханалар құрылды. Онда инструменталды және электро-акустикалық зерттеу жүргізілді [57].

Электро-акустикалық зерттеу әдісі арқылы әуеннің компоненттерін, яғни негізгі тон жиілігінің бағыты мен формасын және өзгеру жылдамдығы мен диапазонын дәл анықтауға мүмкіндік туды. Кейбір ғалымдар интонация модельдерін беруде бұл әдістің тиімсіз екенін атты. Сөйтіп, ағылшын ғалымы Д. Джоунздың әдістемелік құралы жарыққа шығып, интонацияның транскрипциясын беруде графикалық бейнелеу әдісінің тиімді екенін айтып, ұсыныс білдірді [58]. Ғалым тонның жоғарылауы мағынаның маңыздылығын көрсететінін білдіріп, жоғарғы тонның тиянақсыз екенін, ал төменгі тон ойдың аяқталғанын білдіретінін, яғни тиянақты екенін айтады.

Әрі қарай Х. Пальмер интонациялық транскрипцияның жол үсті белгілерін беруді жөн деп санаған. Мысалы, екпін белгісін /' /, әуен көрсеткіші бойынша тонның бағытының жоғарылауы мен төмендеуін көрсету үшін (/↓/ төмендеу /↑/ жоғарылау) белгісін қолданады. Бұл белгілер фразаның аяқталуына қарай сөздің соңына, екпінді сөз алдына қойылады. Ғалымдар барынша интонемаларды дәл беру мақсатында әртүрлі әдістер де қолданғандығы белгілі болып отыр. Тіпті америкалық интонолог ғалым К. Пайк интонеманы, яғни әуен көрсеткішін дәл беру мақсатында санмен белгілеуді жөн көрген. Мысалы, өте жоғары (1), жоғары (2), орта (3) және төмен (3) деп нақты етіп береді.

Қазақ тіл білімінде Ахмет Байтұрсынұлы, Құдайберген Жұбанов 1920 – 1930 жылдары интонация туралы алғаш рет пікір білдіреді. Әрі қарай интонация жайында С. Аманжолов, Н. Сауранбаев, І. Кеңесбаев, Ж. Аралбаев, Ә. Нұрмаханова, З. Базарбаева зерттеу жүргізеді.

Қазақ тіл білімінде Ахмет Байтұрсынұлы интонацияның ерекше сипатын, яғни дауыстың көтеріліп, төмендеп әнше оралып отыратынын «кілемге салған түрдей» деп көрсетеді.

Қазақ тіл білімінде интонацияны арнайы сөз еткен ғалым Ә.Нұрмаханованың еңбегінде интонацияның ерекше көрінісі жан-жақты зерттеліп, интонацияға тән әуені, қарқыны, дауыс ырғағы мен диапазоны, тембрі, екпіні сөз болады [59]. Сондай-ақ академик Р. Сыздықтың еңбегінде сөйлеу мәнерінің өзі біркелкі болмайтындығы, яғни сөздің көтеріңкі, салтанатты үнмен және жай үнмен әртүрлі айтылатындығы сөз болады [60].

Академик З. Базарбаева алғаш рет интонацияның қазақ тіл біліміндегі көрінісін жан-жақты зерттеді. Ғалым интонацияның компоненттеріне әуен, үдемелік, қарқын, кідіріс, ырғақ, екпін, дауыс реңкін жатқызып, қазақ тілінде 8 интонама бар екенін эксперименттік зерттеу барысында анықтады. Олар: тиянақты, тиянақсыз, жалпы сұрақ, арнаулы сұрақ, қатал бұйрық, сыпайы бұйрық, лепті сөйлем, қыстырма сегмент интонамалары [49]. Яғни ғалым осы интонация тәсілдерінің сөйлеу барысында жүзеге асатындығын айтады. Кез келген сөйлемді алсаңыз да просодикалық тәсілдер арқылы жүзеге асатындығын және жағдаятқа қарай өзгеріп отыратындығын байқаймыз.

Интонациялық транскрипциялардың әртүрлі болуы мүмкіндігінше түпнұсқаға жақын сөйленістің моделін қалыптастыруға мүмкіндік береді. Н.Д.Светозарованың еңбегінде «Интонациялық транскрипцияның мақсаты – интонацияны сипаттаудағы теория шеңберінде бірден-бір сөйлеу сегментінің интонациялық суретінің негізгі белгілерін жеткізу» [61] – дейді. Яғни жалпы мәтінді транскрипциялауда фонетикалық және интонациялық транскрипцияны қоса берген жөн. Зерттеуге көркем әдебиет стилінен Ә. Нұршайқовтың «Ақиқат пен аңыз» атты шығармасынан үзінді алынып, оған фонетикалық және интонациялық транскрипция жасалады. Мұнда фонетикалық транскрипция жасауға дыбыстардың акустика-артикуляторлық өзгерісі, сөздердің ассимиляциясы (прогрессивті, регрессивті, тоғыспалы), үндестік заңына сай берілуі жатса, ал интонациялық транскрипцияда сөйлемді синтагмаға, ырғақтық топқа және бунақтық топқа бөле отырып, қазақ тіліне тән 8 интонаманы қою және дауыс үдемелігін, кідіріс пен қарқынын көрсету негізге алынады.

пойызда геле жатқанда → > – | *тайғаның* → | *алдаң ғашқан* > → | *ақ түлкүсүнің құйрұғындай* ↑ < | *жебжеңіл* > → | *жұб-жұмсақ сыйақтанып* – < ↑ | *дала бетінде* – → > | *бұлаңдап тұрған боранның* ↑ < | *жерге түскөнде* → – ↓ *екпіні* → < | *ер жігітті алып соғардай* < ↑ – | *қатт'екен* ↓ || > ↓ | *ауұл ғөрүмбейді* > ↓ | *мынау сансыз* → > | – *ақ түлкүнің құйрұғындай* ↑ < | *бұлаңдаған* ↓ > | *ақ түтөктердің* < ↑ | *ар жағында,* – → | *төртпес шақырым* → | *жерде гана тұр* > ↓ || *жүр-жүрлөп* < ↑ – | *жүрөкті* < → | *сағыныш жетелейді* > ↓ || «*Сабыр* | *сабыр деп* ↑ < – | *санам тартпақтайды* > ↓ | *жас адамға* → < – | *жүрөк* | *бий* > ↓ | *сана бйлейтін* → < | *шар тартқан* → – | *кез емес* → | *еді зой ол* > ↓ || *сақылдаған айаз* → > | *соғып тұрған* → > | *бораңға қарамастан* < ↑ *қайдасың айаулым* – ↑ ↓ *ауулым* < ↑ *қайдасың* < – ↑ | *ат'ана* → *бауұрұм* < ↑ *деп* – < → | *мыңбұлақты бетке алып* – < → | *жүрдүм де геттім* > ↓ [62].

Зерттеуде интонациялық транскрипцины беру мақсатында арнайы белгілер қойылды [63].

Зерттеу барысында мәтіннің интонациялық транскрипциясы аудиторлық талдау негізінде жүргізілді. Сондай-ақ әуені, дауыс үдемелігі, қарқыны, кідірісі анықталып, интонациялық талдау жасалды. Әрбір интонациялық компоненттерінің сөйлемнің мағынасын ашып беруде және ойды нақты жеткізуде рөлі ерекше екендігін дәлелдеу үшін мысалдар келтіріледі. Мысалы:

↑ *пойызда геле жатқанда* → > – | *тайганың* → | *алдаң ғашқан* > → | *ақ тұлкүсүнің құйрұғындай* ↑ < | *жеб-жеңіл* > → | *жұб-жұмсақ сыйақтанып* – < ↑ | *дала бетінде* – → > | *бұлаңдап тұрған боранның* ↑ < | *жерге түскөнде* → – ↓ *екпіні* → < | *ер жігітті алып соғардай* < ↑ – | *қатт'екен* > ↓ ||

Әуен мен үдемелік. Сөз басы тиянақсыз интонемамен басталғанымен, сөз ортасында бірқалыпты интонемамен жалғасып, бірде көтеріліп, соңына қарай бәсеңдей түседі. Әрбір синтагмадағы ой біріне-бірі жалғасып, даланың көрінісін ерекше әсермен беруде просодикалық құралдардың, әсіресе әуен біресе көтеріліп, біресе бірқалыпты айтылатын болса, ал үдемелігі әуенмен қатар жүріп күшейіп, бәсеңдеп отырады.

Қарқыны. Сөйлем басталғанда бірқалыпты қарқынмен айтылып, үшінші синтагмада күшейеді. Бұл сөйлемде қарқынның біресе күшейіп, біресе бәсеңдеуі байқалады. Яғни қарқынның күшеюі сөздің маңызсыз тұстарын көрсетеді.

Кідіріс. Сөйлемде бес кідіріс жасалған. Әрбір кідіріс сөздің мағынасын беру мақсатында жасалады. Сондықтан бұл сөйлемдегі кідірістер логикалық ойды, мағынаны ашып тұр.

↑ *ауұл Әөрүмбейді* > ↓ | *мынау сансыз* → > | – *ақ тұлкүнің құйрұғындай* ↑ < | *бұлаңдаған* ↓ > | – *ақ түтөктердің* < ↑ | *ар жағында* – → | *төртпес шақырым* → | *жерде ғана тұр* > ↓ ||

Әуен мен үдемелік. Бірінші және екінші сөйлемнің басында көтеріңкі әуенмен айтылып, сөз соңында бәсеңдейді. Дауыс үдемелігі әуенмен қатар жүріп, сөз басында күшейіп, соңына қарай бәсеңдеген.

Қарқыны. Бірінші сөйлемде қарқынды айтылып, екінші сөйлем бірқалыпты қарқынмен басталады да, алтыншы синтагмада қарқыны күшейіп, соңына қарай бәсеңдей түскен. Яғни қарқынның бірде күшейіп, бірде бәсеңдеуі сөйлемнің маңызды және маңызсыз тұстарын анықтауға негіз болады.

Кідіріс. Бұл сөйлемдерде үш кідіріс жасалған. Кідірістер физиологиялық және логикалық жағынан келесі айтар ойды жинақтау мақсатында жасалғаны көрінеді.

↑ *жұр-жұрлөп* < ↑ – | *жүрөкті* < → | – *сағыныш жетелейді* > ↓ || *сабыр* | *сабыр деп* ↑ < – | *санам тартпақтайды* > ↓ |

Әуен мен үдемелік. Бірінші және екінші сөйлем лепті сөйлем интонемасымен жасалғандықтан, әуені көтеріңкі леппен айтылған. Сөз ортасынан соңына дейін бірқалыпты әуенмен айтылған.

Қарқыны. Бірінші сөйлемнің басында қарқынды айтылғанымен, сөз ортасынан аяғына дейін қарқыны бәсеңдеді.

Кідіріс. Екі сөйлемде жасалған кідірістер физиологиялық ғана емес, логикалық кідіріске жатады. Өйткені екі сөйлемнің астарында ой мен ерекше сезім жатыр.

жас адамға →[<] – | *жүрек – биі* > ↓ | *сана бйлейтін* →[<] | *шар тартқан* → –
 | *кез емес* → | *еді зой ол* > ↓ || *сақылдаған айаз* → > | *соғып тұрған* → > | *бораңға*
қарамастан < ↑ | *қайдасың айаулым* – ↑ | *ауұлым* <

Әуен мен үдемелік. Екі сөйлемде де бірқалыпты әуенмен басталып, аяғына қарай бәсеңдеген. Дауыс үдемелігі сөз басында бәсеңдеп, сөз ортасында күшейе түседі. Ал үшінші сөйлемнің ерекшелігі сұраулы интонамамен айтылып, көтеріңкі тоналды диапазонмен айтылғандығы көрінеді.

Қарқыны. Бірінші сөйлемде қарқыны бәсеңнен бастап, күшейе түссе, екінші сөйлемде бірқалыптыдан күшейе түскендігі байқалады. Ал үшінші сөйлемде өте баяу айтылады.

Кідіріс. Бұл фразада кідірістің физиологиялық түрінен гөрі, логикалық кідірістің жасалғаны көрінеді.

↑ *қайдасың* < – ↑ | *ата-ана* → *бауұрым* < ↑ | *деп* – < → | *мыңбұлақты бетке алып* – < → | *жүрдүм де геттім* > ↓

Әуен мен үдемелік. Сөз басы сұраулы сөйлем интонамасымен басталғандықтан, көтеріңкі тональды диапазонмен айтылған. Яғни әуені көтеріңкі, ал дауыс үдемелігі бәсеңдей түскен.

Қарқыны. Бұл сөйлемнің басынын бастап, соңына дейін бірқалыпты қарқынмен айтылғандығы көрініп тұр.

Кідіріс. Бұл сөйлемде психологиялық кідіріс жасалған. Өйткені сұраулы, әрі лепті сөйлем интонамасымен айтылғандықтан, сөйлеушінің ішкі көңіл күйі ерекше кідіріс жасау арқылы байқалады.

Қазіргі қазақ тіл білімінде, ауызша сөйлеу кезіндегі орфоэпиялық норма мәселесімен қатар интонацияның да ықпалы зор екендігі академик Р.Сыздықтың, З.Базарбаеваның зерттеулерінде жан-жақты талданған. Мақалада академик З.М. Базарбаеваның зерттеуіндегі интонама (тиянақты, тиянақсыз, жалпы сұрақ, арнаулы сұрақ, лепті, қатал бұйрық, сыпайы бұйрық, қыстырма сегмент) түрлері басшылыққа алынды.

Осы орайда қазіргі «Қазақ тілінің ұлттық корпус» базасында ауызша ішкорпус жұмысын бастады. Онда ауызша сөздің базасы жасалып, мәтіннің просодикалық ерекшелігі көрсетіліп беріледі. Ішкорпус базасында ауызша жазылған мәтіннің дайындықсыз сөйлегендегі орфоэпиясы мен нормаға түскен орфоэпиясы беріледі. Әр сөйлем жеке-жеке синтагмаға бөлініп, оның негізгі просодикалық ерекшелігі, яғни әуенінің белгісі қойылады. Ауызша ішкорпус базасы болашақта ауызша сөз нормасын ажырата алуға, қазақ тілінің үндесім, үйлесім заңдылығын сақтап сөйлеуге көмекші құрал бола алады. Осы ретте сөйлеудің просодикалық ерекшелігін көрсететін әуені (көтеріңкі, бәсең, бірқалыпты) мен дауыс үдемелігі, кідірісі мен қарқыны жан-жақты талданып берілді. Төменде ауызша ішкорпустың қазіргі көрінісі беріліп отыр.

Сұбат беруші	Орфография	Сөйлеуші орфоэпиясы	Бейнефильм	Орфоэпиялық норма
Өзпін Нұршайықов	Менен кейін өкімнің інісі, өкімнің аты Нұршайық, өкімнің інісі Нигат деген кісі, асқарға алады. Бір үйден үшкірші аттанды. Ерекше есте қалғанын, шешем ауыртып жатыр еді. Біз, шана-шана болып, пар-пар аттың ұшын шепті ауылмыздан шыға берген уақытта, ағашында ауылдың іші аралап, әр үйге кіріп-шығып қойлатып, сүйіт еңді мекеттің жанынан жоқта аудан орталығына шыға берген кезде жанғы үйде ауыртып жатқан шешем аяқымыздан шықты.	↑ менен кейін өкімнің інісі → // өкімнің аты нұршайық → // өкімнің інісі нигат деген кісі → // асқарға алады; // ↑ бір үйден үшкірші → // аттанды; // ↑ ерекше есте қалғанын → // шешем ауыртып жатыр еді; // ↑ шана-шана болып → // пар-пар аттың ұшын шепті ауылмыздан шыға берген уақытта → // ағашында ауылдың іші аралап → // әр үйге кіріп-шығып қойлатып → // сүйіт еңді мекеттің жанынан жоқта м. аудан орталығына шыға берген кезде жанғы үйде ауыртып жатқан шешем аяқымыздан шықты; //	↑ менен кейін өкімнің інісі → // өкімнің аты нұршайық → // өкімнің інісі нигат деген кісі → // асқарға алады; // ↑ бір үйден үшкірші → // аттанды; // ↑ ерекше есте қалғанын → // шешем ауыртып жатыр еді; // ↑ шана-шана болып → // пар-пар аттың ұшын шепті ауылмыздан шыға берген уақытта → // ағашында ауылдың іші аралап → // әр үйге кіріп-шығып қойлатып → // сүйіт еңді мекеттің жанынан жоқта м. аудан орталығына шыға берген кезде жанғы үйде ауыртып жатқан шешем аяқымыздан шықты; //	↑ менен кейін өкімнің інісі → // өкімнің аты нұршайық → // өкімнің інісі нигат деген кісі → // асқарға алады; // ↑ бір үйден үшкірші → // аттанды; // ↑ ерекше есте қалғанын → // шешем ауыртып жатыр еді; // ↑ шана-шана болып → // пар-пар аттың ұшын шепті ауылмыздан шыға берген уақытта → // ағашында ауылдың іші аралап → // әр үйге кіріп-шығып қойлатып → // сүйіт еңді мекеттің жанынан жоқта м. аудан орталығына шыға берген кезде жанғы үйде ауыртып жатқан шешем аяқымыздан шықты; //
Өзпін Нұршайықов	Аяқымыздан шыққанда зымпырап келе жатқан екі пар ат, үлкен еңді мықты пар аттар ғой, екеуі де аузымен ыстық тұрған пар ат, зымпырап келе жатқан кезде ана кісі аяқтан шыға келді. Сол кезде мен қорқып ой шешемді талтап кетеді-ау мына аттар деп.	↑ аяқымыздан шыққанда зымпырап келе жатқан екі пар ат → // үлкен еңді мықты пар аттар ғой → // екеуі де аузымен ыстық тұрған пар ат, зымпырап келе жатқан кезде ана кісі аяқтан шыға келді. Сол кезде мен қорқып ой шешемді талтап кетеді-ау мына аттар деп; //	↑ аяқымыздан шыққанда зымпырап келе жатқан екі пар ат → // үлкен еңді мықты пар аттар ғой → // екеуі де аузымен ыстық тұрған пар ат, зымпырап келе жатқан кезде ана кісі аяқтан шыға келді. Сол кезде мен қорқып ой шешемді талтап кетеді-ау мына аттар деп; //	↑ аяқымыздан шыққанда зымпырап келе жатқан екі пар ат → // үлкен еңді мықты пар аттар ғой → // екеуі де аузымен ыстық тұрған пар ат, зымпырап келе жатқан кезде ана кісі аяқтан шыға келді. Сол кезде мен қорқып ой шешемді талтап кетеді-ау мына аттар деп; //
Өзпін Нұршайықов	Негіз өкімнің білмеймін шешем екі қолын жайыл, жанғы аттарды артына тұра қалға уақытта, аттар екеуі бірдей лок етіп тастай қалды. Сөйтіп шешемді баспай аман қалды. Шешемнің айтына: «Қалған сен, собыстан аман келесің, бірде, аман келе ме ағаш келе ме оқи білмеймін, сені келесің жолғы болсын, қалама деді. Менің күш бұтыңа көз алдыма осы сурет келер деді ғұрыда. Қоманда болмады, оқи тостандар осы ғұрыда демалысыңдар деген сөз. Сол жолу-жаулыты жұрп келе жатқан, шыршап келе жатқан адамдар жата-жата қалды әт мен отыра қалып зат қалпымынан Абайды суырып алып, Абайды оқи бастауды қам.	↑ негіз өкімнің білмеймін шешем екі қолын жайыл → // жанғы аттарды артына тұра қалға уақытта → // аттар екеуі бірдей лок етіп тастай қалды → // шешемді баспай аман қалды; // ↑ собыстан аман келесің → // бірде, аман келе ме ағаш келе ме оқи білмеймін → // сен келесің жолғы болсын → // қалама деді; // ↑ менің күш бұтыңа көз алдыма осы сурет келер деді ғұрыда → // қоманда болмады, оқи тостандар → // осырада демалысыңдар деген сөз; // ↑ сол жолу-жаулыты жұрп келе жатқан → // шыршап келе жатқан адамдар жата-жата қалды → // әт мен отыра қалып зат қалпымынан Абайды суырып алып → // Абайды оқи бастауды қам; //	↑ негіз өкімнің білмеймін шешем екі қолын жайыл → // жанғы аттарды артына тұра қалға уақытта → // аттар екеуі бірдей лок етіп тастай қалды → // шешемді баспай аман қалды; // ↑ собыстан аман келесің → // бірде, аман келе ме ағаш келе ме оқи білмеймін → // сен келесің жолғы болсын → // қалама деді; // ↑ менің күш бұтыңа көз алдыма осы сурет келер деді ғұрыда → // қоманда болмады, оқи тостандар → // осырада демалысыңдар деген сөз; // ↑ сол жолу-жаулыты жұрп келе жатқан → // шыршап келе жатқан адамдар жата-жата қалды → // әт мен отыра қалып зат қалпымынан Абайды суырып алып → // Абайды оқи бастауды қам; //	↑ негіз өкімнің білмеймін шешем екі қолын жайыл → // жанғы аттарды артына тұра қалға уақытта → // аттар екеуі бірдей лок етіп тастай қалды → // шешемді баспай аман қалды; // ↑ собыстан аман келесің → // бірде, аман келе ме ағаш келе ме оқи білмеймін → // сен келесің жолғы болсын → // қалама деді; // ↑ менің күш бұтыңа көз алдыма осы сурет келер деді ғұрыда → // қоманда болмады, оқи тостандар → // осырада демалысыңдар деген сөз; // ↑ сол жолу-жаулыты жұрп келе жатқан → // шыршап келе жатқан адамдар жата-жата қалды → // әт мен отыра қалып зат қалпымынан Абайды суырып алып → // Абайды оқи бастауды қам; //

Сурет 3. Ауызша корпустың аудиожазбаның орфография, сөйлеуші және норма орфоэпиясы берілген мәтіндер конкордансы

Қорыта келгенде, фонетикалық транскрипциямен қатар қазақ тілінің интонациялық ерекшелігін беру мақсатында интонациялық транскрипцияның моделі, яғни үлгісі беріледі. Бұл үлгі осы кезге дейін қазақ тіл білімінің супер-сегментті фонетика саласындағы зерттеулердің негізінде қалыптасып, мәтін талдау барысында қолданысқа еніп отыр.

2.7 Орфоэпиялық конвертер жасау тәжірибесі

Сөйлеу барысында қазақ тілінің «сөз сазын бұзып айту» – қазіргі таңда тілші ғалымдар мен ана тілі жанашырларын алаңдататын «тілдің өзіндік ерекшелігін сақтау» мақсаты төңірегіндегі негізгі мәселенің бірі. Айтылуында «Ағнұр, Амаңгелді, жақсы гөрөм, туған гүн т.б.» тізбектеліп кете беретін сөздер мен сөз тіркестері өз сазымен айтылмайды. Қала мектебін, қала берді ауыл мектептерін бітірген түлектердің 80-90 %-ы қазақ тілінің орфоэпиялық нормасын білмейтін (тек теория жүзінде ғана білетін), соны сақтамайтын ұрпақ болып келеді.

Мемлекеттік ұлттық тілдің баршаға ортақтығы осы тілді қолдануда белгілі бір заңдылықтардың, яғни нормалардың болуын талап етеді. Тілдік нормалар оның барлық қаттауында болуы шарт. Норма сондай-ақ тілдің жазба және ауызша түрде қызмет ету барысында да орын алады. Қазақ әдеби тілінің барша заңдылықтарын, оның ішінде ауызша сөйлеу нормаларын бұзбай, дұрыс пайдалануда, тек пайдалану ғана емес, ол нормаларды көпшілікке үйретуде ықпалды тұлғалардың міндеттері мен қызметтері орасан зор. Ата-анадан бастап, балабақша тәрбиешісі, мектеп, университет оқытушылары мен қазақ сахна қайраткерлерінің, радио, теледидар жүргізушілерінің ауызша сөйлеу мәдениетінің жақсаруына ықпал ете алады [60, 121].

Біз басты назарға алып отырған ауызша сөйлеу қағидалары ғылымда орфоэпиялық нормалар деп аталады. **Орфоэпия** (гр. «orthos» – дұрыс, «epos»

– сөз, сөйлеу) – сөздер мен сөз тіркестерінің дұрыс айтылу ережелерінің жиынтығы [60]. Академик Р.Сыздық «Қазақ тілінің анықтағышы» еңбегінде орфоэпияға қатысты «сөз сазы» деген ұғымды қолданады. *Сөз сазы* – сөйлеу үстінде (актісінде) сөздердің дұрыс айтылып, құлаққа жағымды естілуі. Ал *сөздерді дұрыс айту* дегеніміз тілдің табиғи дыбыс заңдылықтарын сақтауы [60, 130]. Дегенмен сөз сазы ұғымын тек орфоэпиялық нормалар жиынтығы деген тар мағынада түсінбеуіміз керек. *Сөз сазы* – ауызша сөздің маңызды құрамдас бөлігі, оны безендірудің, ондағы интеллектуалды ақпаратты жеткізудің, мағынасы мен эмоционалды-экспрессивті реңктерін анықтаудың құралы. Сөз сазының элементтері: тембр, дауыс жоғарылығы мен күші, әуен, қарқын, кідіріс, дауыс модуляциясы, ырғақ, ой екпіні, дикция [65, 37]. Қазақ тілінде сөз бен сөз тіркестері ауызша сөйлеу барысында табиғи түрде үйлесіп, бір-бірімен үндесіп, ыңғайласып айтылуынан олардың жеке тұрғандағы дыбысталуы мен сөйлеу кезіндегі айтылуы әрдайым бірдей болып шықпайды [4, 400]. Қазақ сөзінің сазы туралы айтқанда, ең алдымен үндесім (сингармонизм) құбылысына ерекше тоқталып өту керек. Сөзбен сөз арасында, түбір мен қосымша арасындағы дыбыстардың бір-біріне ықпал етіп, өзгеріп айтылуын үндестік (үндесім) заңы реттейді [65, 37]. Үндесім қазақ тілінің іргелі және ең басты сингармонологиялық заңдылығы болғандықтан, қазақ тілі дыбыстарының үндесім құрамын талдап шығу ғалымдардың басты міндеттерінің бірі болды [66, 104]. Қазақ фонетикасының зерттелім желісі төрт тарихи дәуірден өтті. Осы уақыт бойындағы зерттеулер нәтижесінде қазақ тілі дыбыс құрамы айқын, төл дыбыстары сараланып, үндесімнің 4 әуезі ескерілген, үндесім/үйлесім дыбыстар құрамы анықталған тілге айналды [66, 29]. Сондай-ақ ғылым-техниканың дамуымен фонетиканың жаңа бағыттары бойынша зерттеулердің толықтырыла берері анық.

Әр тарихи кезеңде шындалып, тұрақталған ауызекі сөйлеу тілінің нормасына бүгінгі таңда үлкен міндет пен талап қойылуда. Демек, оқушылар қауымы үшін емле ережесі мен грамматика ережесін білу қаншалықты қажет болса, сөздің дыбысталу заңдылығын меңгерудің маңызы да соншалық. Сондықтан ана тілінің сөйлеу нормасын саналы түрде үйрету қазіргі таңдағы ең актуалды мәселе [67, 402]. Осы аталған актуалды мәселе тілші ғалымдарды да толғандырады әрі бірден-бір мәселені шешуші тұлғалар деп есептей аламыз.

2003 жылы филология ғылымдарының докторы Н. Уәлиұлы «Біз қалай сөйлейміз?» атты бағдарламада: «Орфоэпиялық нормаларды жөндеудің бір жағы тілші ғалымдардың халық қолданысы үшін сөздіктер, әртүрлі анықтағыштар әзірлеу қажеттігіне байланысты. Бұған қатысты іргелі зерттеулер жүргізіліп, көптеген еңбектер баспадан шығарылып та жатыр. Екінші жағы осы шығарылған еңбектердің БАҚ-та насихатталуы жолға қойылуы керек» [68, 64] деген пікір білдірген. Біздің назарға алып отырған нысанымыз ауызша тіл мәселесі болғандықтан, жарық көрген орфоэпиялық сөздіктер мен анықтағыштарға тоқталсақ. Алғашқы орфоэпиялық сөздік Қ.Нетәлиеваның «Қазақ тілінің орфоэпиялық сөздігі» (1977), М. Дүйсебекова «Қазақ тілінің қысқаша орфоэпиялық сөздігі» (1981), Р.Сыздық «Сөз сазы» (1983, 1995), Ә.Жүнісбектің басшылығымен шыққан «Қазақ тілінің орфоэпиялық сөздігі»

және соңғы Ғ.Әнес басшылығымен шыққан «Орфоэпиялық сөздік» (2007) құрастырушылары сөздікті түзу барысында қазақ тіл білімінің фонетика, фонология, интонология, сөз мәдениеті т.б. салалары бойынша соңғы зерттеулердің нәтижесін негізге алып құрастырған [69, 17]. Ал қазақ тілінің орфоэпиялық анықтағышы 2000 жылы шыққан Р.Сыздықтың «Қазақ тілінің анықтағышы» және Н.Уәлидің басшылығымен 2004 жылы шыққан анықтағыш. Қазақ тілінің орфоэпиялық нормалары бекітіліп, негізгі ерекшеліктері аға ғалымдардың еңбектерінде зерттеліп, тиянақталған. Ендігі мәселе – сол еңбектердің халық қолданысына заманауи техниканың мүмкіндігін пайдалана отырып, интерактивті құрал ретінде ұсынылуы.

Қазір дүниежүзінің дамыған мемлекеттері VI технологиялық укладты қалыптастыруда. Олар: нано-биотехнология, гендік инженерия, информациялық-коммуникациялық технология, космостық технология. Бұл салаларда біздің үлесіміз жоқтың қасы. Жалпы, біз технологиялық укладтан 40-50 жыл артта қалып отырмыз [70, 180]. Біз ғылымның жаңалықтарын, компьютер тілін, жаңа интернет технологияларды, программамен қамтамасыз етуді, роботтандыруды, цифрлы экономика тілін тез игермесек, тіпті қалып қоятын түріміз бар. КСРО құрамында болған Эстония қазіргі таңда интернет жылдамдығы ең жоғарғы ел. Бұл ел де Сингапур мемлекеті сияқты ең көп қаржыны бастауыш мектеп біліміне салады екен. Бағдарлама жасаудың негіздерін бірінші сыныптан бастап оқытады, компьютер тілі оларға француз тілінен артық. Эстония біз сияқты саны аз халықты инновацияға тез икемдеп, дамытуға болатынын көрсетті. Егер бұрын өндіріске керегі темір, ауыл шаруашылығына жер болса, қазір цифрлы дәуір келді. Экономикамыз информацияға негізделіп отыр. Алда кім, қай ел тез арада үлкен ақпараттарды, интернет жүйесін бірлестіктерді, жасанды интеллекті т.б. игерсе, солар озады, әлемде алда болады [70, 257]. Қазақ тіл білімінің қай саласын алсақ та, құнды ақпаратқа толы зерттеулерге куә боламыз. Қолғанат (қолданбалы) лингвистердің мақсаты – қазақ ғылымында әлдеқашан зерттеліп қойған білімдерді халық игілігіне, қолданысына даярлап беру; ғалымдарымыздың ерен еңбегін заман талабына сай жасақтап, халыққа ұсыну; қазақ тіл білімінің майталман ғалымдарының көз майын тауысып жасаған еңбектерінің халық қолданысына енбей тұншығып жатуы жанымызды қынжылтады. Ендігі кезекте қолғанат (қолданбалы) лингвистер мен бағдарламашылар бірлесе отырып, қоғам қажеттілігін өтеуге тиіс. Ахмет Байтұрсынұлы атындағы Тіл білімі институты қазақ сөз сазының өзіндік ерекшеліктерін сақтап сөйлеуді қамтамасыз ету үшін орфоэпиялық конвертер әзірледі.

Конвертер – бір типтегі файлдарды басқа түрдегі файлдарға түрлендіретін арнайы бағдарлама [71]. **Орфоэпиялық конвертер** – орфографиямен жазылған сөзді, сөз тіркесті, сөйлем мен мәтінді орфоэпиялық нормаға сай жазып беретін бағдарлама. Орфоэпиялық конвертер жасау барысында байқағанымыз, анықтағыштың қолданысқа ауадай қажет екенін ұғындық. Сондай-ақ анықтағыштар мен сөздіктердің тек кітап түрінде ғана емес e-book (интерактивті кітап) ретінде де қолданысқа ұсынылып, сол еңбектер негізінде арнайы орфоэпиялық нормаларды меңгеруге арналған қосымшаның да жасалуы маңызды.

Орфоэпиялық конвертер жасалу барысында осыған дейінгі қазақ тілінің орфоэпиялық нормалары зерттелген негізгі еңбектер басшылыққа алынды. Сол еңбектерге қысқаша тоқталсақ.

Академик Р.Сыздық «Қазақ тілінің анықтағышы» еңбегінде орфоэпиялық ережелерді 36 бапқа (параграфқа) бөліп, әрбір ереженің түсіндірмесін, дәлелдемесін қоса ұсынған. Сонымен қатар ережені бекіту үшін тапсырмалар беріліп, мысалдардың жазылуы да жиі көрсетіліп отырады. **10-бапта:** «Сөздің басында немесе бірінші ашық буындағы **ұ, ү, о, ө** дыбыстары және екі дауыстының ортасында келген **у** дыбысы (тауық, бауыр) келесі буындағы **ы, і** дыбыстарын өзгертіп, **ұ, ү** дыбыстарына жуықтатып естірмеді. Мысалы, құлын-құлұн, бүгін-бүгүн» ережесі берілген. Дәлелдеме ретінде көптеген мысалдарды келтіре отырып, өлең шумағын дұрыс айтып үйренуді ұсынады.

*Қазақтың / өзгөжұрттан / сөзү / ұзұн /
Бірінің / бірішатшаң / ұқпассөзүн //
Көздүңжасы / жүрөктүң / қаныменен /
Ерітуге / болмайды / ішкімүзұн //*
(Абай) [60, 167]

Берілген 36 бап қазақ сөз сазының жеке дыбыстар, сөз іші, біріккен сөз, тіркескен сөз, кірме сөз ережелерінің негізгі заңдылықтарын анықтап, түсіндіріп береді. Бағдарламалау тілінде айтсақ, негізгі 36 алгоритім әзірленген.

Келесі еңбек – Н.Уәлиұлы, Ғ.Әнес, Қ.Күдерінова, А.Ғазылжанның құрастыруымен 2004 жылы радио, -тележурналистерге көмекші құрал ретінде жарық көрген анықтағыш. Анықтағышта ауызша сөйлеу мәдениеті, маңызы сөз болып, зерттеулер негізінде қысқаша орфоэпиялық сөздік жобасы ұсынылған. Сөздіктің құрамы мен құрлысын түсіндіре келе, жеке дыбыстардың 38 алгоритімі, сөз ішінде 42, тіркес арасында 29, жалпы 109 алгоритім жүйеленіп берілген. Осы жоба негізінде 2007 жылы 28 000-нан аса сөзді қамтыған орфоэпиялық сөздік жарыққа шықты. Жасалып жатқан орфоэпиялық конвертерге аталған алгоритімдер түгелімен енгізілді.

Автоматтанған орфоэпиялық конвертер жасау идеясы 2022 жылы филология ғылымдарының докторы А.Жаңабекованың жетекшілігімен қолға алынды. Институттың қолданбалы лингвистика және фонетика бөлімі аталған іске жауапты болды. Осыған дейінгі фонетикалық еңбектерді зерттей келе, толықтырылған орфоэпиялық ережелердің алгоритімін құрды. Жоғарыда аталған екі еңбекпен қатар профессор Ә. Жүнісбек еңбектері басшылыққа алына отырып, осыған дейінгі зерттеулер сараланып, жинақталып, толықтырулар енгізу жұмыстары жүрді. Нәтижесінде 14 дыбыстың (и, ю, я, у, щ, л, н, б, с, т, п, қ, к, з) өзіне тән ережелері толық қамтылды. Тіркескен сөздер мен біріккен сөздердің 84 алгоритімі құрылды. Бір ғана и дыбысының алгоритімін мысалға келтірсек:

И дыбысы.

1. Егер сөз жуан тембр: (сөз құрамында - **а, ы, ұ, о**).

1.1. Егер сөз жуан езулік тембр (сөз құрамында - **а, ы**):

-дауыссыздан кейін, онда **ЫЙ (қырғИ – қырғЫЙ)**;

-сөз басында, онда **ЫЙ (Иық - ЫЙық)**

1.2 Егер сөз жуан еріндік тембр (сөз құрамында - ұ, о), онда ҰЙ (оқИ – оқҰЙ)

2. Егер сөз жіңішке тембр (сөз құрамында - ә, і, е, ү, ө).

2.1 Егер сөз жіңішке езулік тембр (сөз құрамында – ә, і, е)

-дауыссыздан кейін, онда ІЙ (кИ – кІЙ);

-сөз басында, онда ІЙ (Ине – ІЙне)

2.2 Егер сөз жіңішке еріндік тембр (ү, ө), онда ҮЙ (үсИ – үсҮЙ)

3. И дыбысы Я дыбысының алдында:

3.1 Егер сөз жуан езулік тембр (а, ы), онда Ы (сИЯ – сЫЯ)

3.2 Егер сөз жіңішке езулік тембр (ә, і,), онда І (әИЯ – әІЙ)

3.3 Егер сөз жуан еріндік тембр (о, ұ), онда Ұ (бұртия-бұртұя)

3.4 Егер сөз жіңішке еріндік тембр (ө, ү), онда Ү (дүрия-дүрүя)

4. И дыбысы Ю дыбысының алдында:

4.1 Егер сөз жуан езулік тембр (а, ы), онда Ы (кИЮ – кЫЮ)

4.2 Егер сөз жіңішке езулік тембр (ә, і), онда І (кИЮ – кІЮ)

4.3 Егер сөз жуан еріндік тембр (о, ұ), онда Ұ (қонжю-қонжұю)

4.4 Егер сөз жіңішке еріндік тембр (ө, ү), онда Ү (үрпю-үрпүю)

Ал төменде жоғарыдағы мысалдың автоматқа енгізілетін бағдарламалық кодын ұсынамыз.

If i.upper() == u'И':

if prev_char == _None:

if main_timbr in vowel_hard_nonlabials: curr_char = u'ЫЙ'

if main_timbr in vowel_hard_labials: curr_char = u'ҰЙ'

if main_timbr in vowel_soft_nonlabials: curr_char = u'ІЙ'

if main_timbr in vowel_soft_labials: curr_char = u'ҮЙ'

if prev_char in consonants:

if cont_timbr in vowel_hard_nonlabials: curr_char = u'ЫЙ'

if cont_timbr in vowel_hard_labials: curr_char = u'ҰЙ'

if cont_timbr in vowel_soft_nonlabials: curr_char = u'ІЙ'

if cont_timbr in vowel_soft_labials: curr_char = u'ҮЙ'

if prev_char == u'y':

if cont_timbr in vowel_hard_nonlabials: curr_char = u'ЫЙ'

if cont_timbr in vowel_soft_nonlabials: curr_char = u'ІЙ'

if curr_char == i:

if prev_char == u'б': curr_char = u'ІЙ'

if prev_char == u'ж': curr_char = u'ЫЙ'

if prev_char == u'з': curr_char = u'ІЙ'

if prev_char == u'з': curr_char = u'ЫЙ'

if prev_char == u'к': curr_char = u'ІЙ'

if prev_char == u'қ': curr_char = u'ЫЙ'

if prev_char == u'с': curr_char = u'ЫЙ'

if prev_char == u'т': curr_char = u'ІЙ'

if prev_char == u'ш': curr_char = u'ІЙ'

if next_char == u'я':

if last_timbr in vowel_hard_timbr: curr_char = u'Ы'

```

if last_timbr in vowel_soft_timbr: curr_char = u'Г'
if next_char == u'ю':
if last_timbr in vowel_hard_timbr: curr_char = u'Ы'
if last_timbr in vowel_soft_timbr: curr_char = u'Г'

```

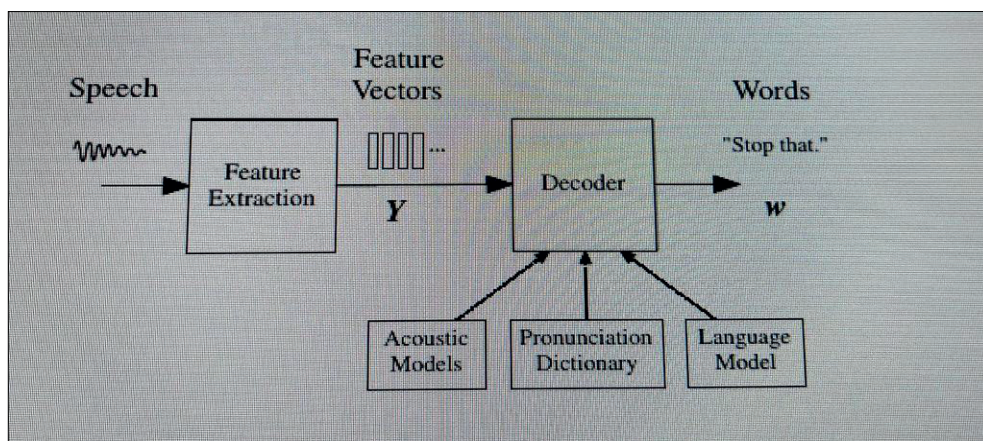
Автомат конвертер жасауға фонетист ғалымдар Ж.Жұмабаева, А.Аманбаева белсене ат салысты. Жалпы саны 183 алгоритмді нақты мысалдармен дәлелдей отырып, бекітті. Сондай-ақ қолданбалы лингвистика бөлім қызметкерлері аға бағдарламашы Ғ.Аязбаев ұсынылған алгоритмді **Python** бағдарламалау тілі арқылы кодтап, конвертердің автоматтануын жүзеге асырды. Бөлім меңгерушісі А.Жаңабекова кездескен қайшылықтарды түзетуде қос сөз, біріккен сөз реестрімен қамтамасыз етті. Осы жұмыстың барысын тікелей қадағалап, жұмыстың сапалы нәтижемен аяқталуын бастан-аяқ жүзеге асырды. Аға лаборант С.Бейсенхан автомат конвертердің сапасын, жіберілген қателіктер мен қайшылықтарды анықтау мақсатында орфоэпиялық сөздікте қамтылған 23 000-нан аса сөздің excel-де салыстырылған нәтижесіне талдау жасады (4-сурет).

Орфография	Конвертер	Верно	Орфоэпия
аба	аба	ИСТИНА	аба
абадан, абаданға	абадан, абаданға	ИСТИНА	абадан, абаданға
абажа	абажа	ИСТИНА	абажа
абажадай	абажадай	ИСТИНА	абажадай
абай, абай болу	абай, абай бол^ү</sup>У	ложь	абай, абай болуу
абай-қоқай	абай-қоқай	ложь	абай-қоқ ай
абайлағыш	абайлағыш	ИСТИНА	абайлағыш
абайлату, абайлатса	абайлатУУ, абайлатІса	ложь	абайлатуу, абай латса
абайлау	абайлаУ	ИСТИНА	абайлау
абайсыз	абайсыз	ИСТИНА	абайсыз
абайсызда	абайсызда	ИСТИНА	абайсызда
абайсыздық, абайсыздығы	абайсыздық, абайсыздығы	ИСТИНА	абайсыздық, абайсыздығы
абайтану	абайтанУУ	ИСТИНА	абайтануу
абайтанушы	абайтанУУшҰ	ИСТИНА	абайтануушү
абақ (күрал), абағы	абақ (күрал), абағы	ложь	абақ, абағы
абақты	абақты	ИСТИНА	абақты

Сурет 4. Автомат конвертер мен орфоэпиялық сөздіктің салыстырылуы

Нәтижесінде жаңадан 10 алгоритім енгізілді. Сондай-ақ біріккен, кірме, термин сөздер жеке-жеке реестр арқылы берілген жағдайда, сол сөздердің орфоэпиялық дұрыстығын қамтамасыз етуге болатыны айқындалды.

Жасалған орфоэпиялық конвертер бірнеше маңызды мәселені шешуге қауқарлы. Біріншіден, ҚТҰК-ның (<https://qazcorpus.kz/>) ішкі белгіленімін толықтырып, сөздерді буынға дұрыс бөледі. Екіншіден, ауызша корпус базасына енетін мәтіндерді орфоэпиялық нормаға сай жазып береді. Үшіншіден, жеке өнім ретінде қолданысқа ұсынылады. Орфоэпиялық конвертерді ғылыми зерттеу жұмыстарында, тілді оқыту мен үйренуде, күнделікті қажет болған жағдайда т.б. жағдайларда пайдалана аламыз. Сонымен қатар, сөйлеу технологияларын жасауда автоматтанған орфоэпиялық нормалардың, сөздіктердің болуы сапалы нәтиже үшін қажет [72].



Сурет 5. Сөйлеуді мәтінге айналдыру компоненттері

Профессор Н.Уәли бір сұхбатында: «Алдымен біздер, мысалы, тіл мамандары жұртшылықты тиісті құралдармен қамтамасыз етсек, ал жалпы қауым өз сөзіне қарауыл қойып сөйлесек» [68, 69] деп түйіндепті. Біздің мақаламыздың түйіні де осы ойға саяды. Тілші ғалымдар тиісті құралдармен қамтамасыз етуде. Ендігі кезек әрбіріміздің өз еншімізде.

3 ОНОМАСТИКАЛЫҚ ІШКОРПУС

3.1 Ономастикалық корпус: антропонимдер мен топонимдердің мәдени тілдік кодын ашудың ғылыми негіздері

Қазіргі жаһаният заманындағы үрдісте адамның, ұлттың өзін-өзі тануы қоғамдық санада үнемі жаңғырып отыратын өзекті мәселелердің біріне айналып отыр. Ұлттың рухани болмысын тіл дүниеліктерімен де байланысты танытын антропоцентристік бағыттағы парадигмада ономастикалық зерттеулердің ерекше орын алатыны сондықтан. Елдің кім екенін, жер-судың кімдікі екенін, кімнің кім екенін көп жағдайда ономастикалық атаулар айғақтап тұрады. Осы кезге дейін ономастикалық атаулардың, оның қатарында антропонимдердің тілдік сипаттамасы жасалып, едәуір нәтижелерге қол жетті. Ендігі тұста тілдің ономастикалық қорындағы мұраларды мәдени семантикалық қырдан қарастыру, оның тереңірек қабаттарындағы кодқа салынған ақпараттарды айқындай отырып, ономастикалық корпусстың базасын күшейту бүгінгі күннің өзекті мәселесі.

Қазақ антропонимдері мен халықтық топонимдердің мазмұн межесінде кездесетін мәдени-тарихи әрі танымдық ақпараттық қабаттар әлі де ашылмай жатқан кеніш іспетті. Байырғы жер-су, елді мекен атаулары, бір есептен, халықтың тарихи-мәдени санасында географиялық карта іспетті деуге болады. Онда, оқи білсек, физикалық географияға да, саяси-экономикалық географияға да тән ақпараттық мазмұн жатыр.

Қазақтың төрт түлік өсіріп, мал өнімін өндірушілік ата кәсібі жер бедері, оның ойы мен қыры, оты мен суы, шөбінің ащы-тұщысы, күн райы, аспан шырақтары жайында эмпирикалық кешенді рационалды білімдер жүйесін қалыптастырды. Бұлардың қатарында эзотериялық наным-сеніммен байланысты мифтік таным-білімдер де ұшырасады.

Этностың тұрмыс-тіршілігімен, табиғат объектілері, дәстүрлі мал шаруашылығы, егіншілік мәдениеті, өндіріс өнімдері, әлеуметтік қатынастар жүйесіндегі, этикалық-моральдық құндылықтар, билік пен құқық саласындағы стереотиптер, діни, мифологиялық наным-сенімдермен байланысты мәдени-тілдік ұжымның санасындағы **ұғым, түсінік, концепті** түрінде орныққан ойтанымдық түзілімдер ерекше орын алады. Бұндай менталды түзілімдердің тілдік референттері **атаулар, фразеологизмдер, нақыл сөздер, мақал-мәтелдер, күрделі фразалық бірліктер** болып табылады.

Осы айтылғандарға сәйкес этнос санасында ақиқат дүние менталды және тілдік бірліктерден құрылымдалған түзілімдер зерттеу **мәдени семантикалық компонент** деп аталды.

Тек қана тілдік семантика тұрғысынан қарағанда *балта*, *балға*, *орақ* сөздері «еңбек құралдары» деген лексикалық мағыналардың референттері болып табылады. Сөздікте *балта* «өткір жүзді, екінші жағы шүйделі, шабуға, қидалауға арналған сабы бар құрал» (Қазақ сөздігі, 186-б.) деп лексикографияланған. Ал балаға ат қойған кезде атаудың құрамындағы шүйделі, сабы, шабу, қидалау деген мәнділіктер бейтараптанып, «өткір», «өткір жүзді» деген мәнділіктер ба-

сым сипат алып, құралдың ұстаған қолға күш беретін ерекше қасиеті айрықша мән алады.

Ономастикалық корпустағы кейбір қазақ антропонимдерінің мәдени семантикасының берілуі. Байырғы қазақ қоғамында материалдық игілікке ие аса дәулетті жіктің өкілі *бай* деп аталады. Қазақ байлары алтын, күміс, інжу, маржан, асыл тастардан гөрі «алтын-күміс тас екен» деп төрт түлікті аса жоғары қояды. Өйткені төрт түлік жыл сайын төлімен көбейіп тұрады. Ал алтын-күміс, асыл тастар қазақы байдың байырғы түсінігінде өздігінен көбеймейтін сандық түбіндегі өлі байлыққа жатады. Оларды «қозғалысқа» салу үшін біреуден арзанға алып, келесі біреуге қымбатқа сатуды еңбекпен, маңдай термен тапқан дәулет деп тани қоймайды.

Байырғы дәстүрлі дала салтында байларды *жылқылы бай, түйелі бай, қойлы бай* деп бөлген. Байлықтың эталондық мөлшері жылқы мен қой түлігі үшін мыңмен, ал түйе түлігі үшін жүзбен өлшенген.

*Мың жылқылы байлардан несі кейін,
Оразасын үйінен ашқан кісі, –*

дейді байырғы жырларда. Түйенің саны жүзден асқан мал иесі ырымдап, түйесінің бір көзін ағызады екен. Қазақ фольклорында «Соқыр түйе» аталатын ертегі бар. Мал иесінің «*жүз түйем бар демей түйеден бір соқырым бар*» дейтіні тіл-көзден сақтанып, келе-келе түйесінің көптігін эвфемизммен тасалайды.

Халықтық антропонимдерде жылқы, түйе, қой атауларының белсенді жұмсалыуының детерминациясы «*алалы жылқы, ақтылы қойға өріс толсын, келе-келе түйе айдасын*» деп, ырымдап қазақы ортада балаларының аттарын *Жылқыбай, Қойлыбай, Қойшыбай, Түйебай* қоятыны белгілі, бірақ бұндай материалдық игілікпен байланысты есімдер себепсізден-себепсіз қойыла салмайды. Бие құлындап, түйе боталап, қой қоздап жатқан кезде туған нәрестенің ырысы мен риздығы болсын деп ниет етеді.

Ал қошқарға байланған күйекті шешіп, қойға салар кездің малшы ауылда өз ырымы бар. Ол ырымды байдың бәйбішесі атқарып, «сідігің алтын, сүргенің саулық болсын» дейді. Сол кездері ауылдарындағы келін босанып, ұл туса, «мықты», «күшті болсын» деп, атын *Қошқарбай* қояды. Әдетте мал танитын білікті шопанның айтуымен күйек таңдаулы, айрықша белгілері бар қошқарға байланады, ал «сүргенің саулық болсын» дейтіні, қой тұяғы саулықтан көбейеді.

Сондай-ақ қойдың төл төгіні – жаппай қоздап жатқан кезде – туған нәрестенің атын *Қозыбай, Қозыбақ* қояды.

Азды-көпті мән берсек, кісі есіміндегі бұндай мәдени-семантикалық ақпараттар этностың дәстүрлі мал шаруашылығының, оның әлеуметтік өміріндегі алатын орнының басым бағытта болғанына байланысты.

Кісі есімдерінің төрт түлікке қатысты әуелдегі мазмұн-межесіндегі кодқа салынған ақпараттар, жоғарыда айтылғандай, үнемі ашық-айқын бола бермейді. Дәстүрлі антропонимикада төрт түлікпен байланысты, бірақ ішкі мазмұн

межесі бүркемеленген, тасаланған есімдер аз кездеспейді. Өйткені кодқа салынған ақпараттар қазіргі тіл иелмендері үшін көмескіленген немесе мүлде жабық.

Мысалы, *Ақбай, Қарабай, Байғара/Байқара* сияқты дәстүрлі есімдердің мазмұн құрылымындағы кодқа салынған танымдық ақпараттар көмескіленіп кеткен. Аталмыш есімдердегі *ақ, қара* компоненті нені білдіреді? Бұлардың детерминанттарын әдетте түр-түспен, онда да олардың туынды мағыналарымен байланыстырып, *ақ* дегенді «адал, ақ сенім», *қараны* «ірі», «үлкен» дегенмен немесе «қара ниет» ұғымымен ұштастыра қараушылар бар.

Біз аталмыш есімдердің пайда болу себептерін (детерминанттарын) құндылықтар немесе антиқұндылықтар емес, атап айтқанда төрт түлікпен байланысты қараймыз.

Қазақ тілінің байырғы сөздік қорында жылқыны *ақ мал* деп атаған. Объектінің бұлай айтылу себебі *жылқыда от жоқ*, жусауы мен жайылым өрісі өзге түліктен өзгеше. Жылқының жейтін шөбін малшы қауым *ақ от*, кейде *боз от* деп атаған. Ақ отқа немесе боз отқа, әсіресе *бетеге*, сол сияқты *көде, ебелек, тарлау* сияқты т.б. қырдың теріскей бетіндегі шөбі жатады. Қазақтың ата сөздерінде бетегелі белге бітем, найза ұстаған ерге бітем; дүниеде не тәтті тарлау тәтті, тарлау жеген жылқының еті тәтті; ебелек отқа семірген дейді. Жылқы түлігінің байырғы кездері ақ мал деп аталуы оттайтын отының ақ от екеніне байланысты. Байқауымызша, халықтық ономастикадағы Ақбай, Ағыбай сияқты есімдердегі ақ атауының мазмұн межесінде «жылқы түлігін» меңзейтін ақпарат кодқа салынған.

Ал, салыстыра кетсек, Ақмырза, Баймырза, Жақсыбай, Мырзаберген, Жомарт сияқты т.б. есімдердің мазмұн межесінде отбасындағы, қоғамдағы рухани құндылықтарды жоғары қойып, ізгілікті дәріптеумен байланысты этикалық ақпарат кодқа салынған.

Халықтық ономастикадағы дәстүрлі антропонимдерге жататын *Қарабай, Байқара/Байғара* атауындағы қара компонентінің мазмұн межесінде кодқа салынған ақпараттарды да төрт түлікпен байланысты ашамыз. Қарабай, Байқара атауларындағы қара компоненті түйе мен қой түлігіне қатысты. Түйе мен қойды, сиырды ежелгі малшы қауым күйіс қайыратын қара малға жатқызады. Түйе мен сиырды ірі қара, ал қой-ешкіні уақ мал деп ажырата қараған. Дәстүрлі қазақ есімдеріндегі Қарабай, Байқара дегендегі қара түйелі, қойлы байлардың антропонимдік «синонимдері».

Қара мал түп төркіні түйе мен қойдың отайтын отына байланысты. Түйе мен қойдың өрісі – қара отты, ащы шөпті жерлер. Содан қара мал атанған. Ал қара отқа халықтық ботаникада жантақ, жусан, изен, теріскен, қарабарак, қоянжын, майқара, көкпек, түйеқарын, түйекұйрық, түйесораң тәрізді т.б. түздің ащы шөптері жатады. Аталмыш қыр шөптерінің қара деген түр-түспен аталуы себебі бұлар қырқалы, қыраттың қар көп жатпайтын күнгей беткейінде, далалы аймақтың сортаң, шөлейт жерлерін бауыр басып өседі.

Әдетте қырды, қырканың күнгей жағы қыстаулық жер болады. Онда қар көп жатпайды, жайылымдық жері қарайып, қара-қоңыр тартып тұрады. Жетіқара, Үшқара, Желдіқара дегендердің қыстау екенін малшы қауым атынан-

ақ айтқызбай-ақ жазбай таниды. Малшы ауылдың күзеуден қыстауға көшкен шағында туған бала Қыстаубай деп аталған.

Сонымен малшы қауымның тәжірибеден туындаған рационалдық таным-білімі түз шөбі, жайылымдық жердің теріскей, күнгейі; күнгей шөбінің қара от, теріскей беттің ақ от (боз от); шөпкіндікті малдың ақ мал, қара мал аталуы түліктің оттайтын отына байланысты аталған. Бұл айтылғанды пысықтай түссек: жылқының жайылымы – ақ от, сондықтан ақ мал аталған; түйе мен қойдың жайылымы – қара от, сондықтан қара мал аталған.

Ал субъектінің (мал иесінің) Ақбай аталуы «ақ малға бай болсын» деген тілек мәнді этикалық ұғымды аңғартады. Сол сияқты күнгейдің қара отына кіндігі байланған түйе, қой түлігі қара мал аталған. Соған орайлас субъектінің Қарабай, Байқара деп «паспортталуының» (онимделуінің) уәжі «қара малы (қой, түйесі) мыңғырған бай болсын» деген тілекті білдіретін этикалық танымға сәйкес ізгі ниеттен туындағаны кәдік келтірмейді.

Малшының дүниетанымдық кеңістігіндегі табиғат объектілері – бір-бірімен байланысты түзілімдер. Қырат-қырқалы жерлер, олардың күнгей, теріскей беттері, ондағы жайылымдық өрістер, қара малдың жайылымы; ақ малдың жайылым өрісі, күнгейдегі қара оттың, теріскейдегі ақ оттың шығымы т.б. бір-бірімен байланысты рационалды танымдық түзілімдер, сайып келгенде, этнос санасындағы ақиқат дүниенің бейнесі когнитивтік әрекеттің барысында Ақбай, Қарабай, Байқара болып онимделген («паспортталған»).

Малшы қауымның өмірден түйген рационалды танымындағы білімінде табиғат объектілерінің – жер бедері, жер оты, өрістегі түліктер өзара байланыстағы үзіктер (сегменттер) когнитивтік түзілімде Ақбай, Қарабай, Байғара сынды халықтық антропонимдердің онимделуінің детерминациясы болып табылады.

Этностың дәстүрлі малшылығымен байланысты қазақ тілінің сөздік қорында әралуан тұрақты сөз орамдарының біразы бай, байлық; кедей, кедейлік ұғымдарымен байланысты қоғамның әлеуметтік сипатын танытады. Айталық, қара қазан бай, шығай бай, қара қасқа кедей, қу кедей, шіріген бай, сіңірі шыққан қу кедей, қарын бай (түк татырмас сараң бай) т.б.

Қара қасқа кедей дегендегі қасқа – жылқының түр-түсіне қатысты «менмұндалап» тұратын ерекше сырт белгісі. Соған сәйкестендіріп, нағыз кедейді қара қасқа кедей дейді. Осы айтылған қара қасқа кедейдің байығанын қара қазан бай деп те айтады. Байлыққа енді ғана қолы жеткен қара қазан бай тумысында өте сараң болады. Қазақ халқының ежелгі фольклорындағы атақты Шықбермес Шығайбай – қара қасқа кедейден байыған мал иесінің типтенген бейнесі.

Ал Шығайбай дегендегі шығай «кедей» деген мағынаны білдіреді. «Орхон-Енисейдегі» Күлтегін ескерткішінде «Чығай жұртты бай қылдым» дейді.

Этностың аксиологиялық танымында қас кедейден байыған қара қазан байлар немесе шығай байлар жағымсыз бейнеде көрінеді. Оны халық мақалында айтылатын «кедейді байыды деме, маңдайының арылмайтын соры бар, байды кедей болды деме, сабасының сарқылмайтын қоры бар» деген этикалық пайымдауы біреу байып, біреу кедейленіп жататын тұстағы әлеуметтік үрдісті

айқын аңғарта отырып, дүниауи (материалдық) құндылықтарды терістемейтіні халықтық антропонимдер жүйесінде «бай» компонентінің белсенді әрі жиі жұмсалып үрдісінен байқалады.

Дәстүрлі антропонимдер жүйесінде шығай байлар мен қара қазан байларға құндылық және антиқұндылық дихатомиясы бойынша сары байлар қарама-қарсы қойылады.

Жеті тұстан жеті атасынан бері байлығы үзілмейтін, еш уақыт қара сабасы ортаймайтын, үйірлі жылқы айдаған байды Сарыбай деп, сары атауымен байланыстыра айтады. Өйткені сары ахроматизмі «ұзақ» деген ұғымды білдіреді. Сары уайым, сары аяз, сар(ы) желіс дегендердегі сары компонентінің мазмұн межесінде «ұзақ», «үздіксіз», «көпке дейін созылған» деген семантикалық элемент (сема) бар. Ойтанымдық (когнитивтік) әрекеттің (актінің) барысында антропонимдік атаудың тілдік матрицасы болып қалыптасқан. Диахрондық тұрғыдан сары байлар «әуелден бай», «үйір-үйірімен жылқы айдаған байлар», ал синхрондық тұрғыда «әйтеуір байлар» деген әлеуметтік жіктің өкілі дегенді меңзейтіні – үстірт ақпарат.

Халықтық ономастикада Ақтоғай, Сартоғай топонимдермен кіндіктес антропонимдер де жиі кездесіп отырады: Байтоғай, Тоғай, Тоғайбек т.б. Бұндағы тоғай компоненті топонимдерге де, антропонимдерге де ортақ болуы назар аудартпай қоймайды. Олардағы ортақтастықтың себебін (детерминатын) айқындай түсуге тура келеді.

Тілдік санадағы тоғай атауы «өзен-су бойындағы бұта аралас өскен әртүрлі ну ағаш» деген географиялық объектіге тән ұғымды білдіреді. Алайда малшы қауымның танымындағы тоғай ұғымының мазмұндық құрылымы бұл айтылғаннан әлдеқайда кең. Атап айтқанда, тоғай ұғымы «қолда, қоражайда бағылмайтын ақ малдың (жылқы түлігінің) қысы-жазы оты мен суы әрі алты ай жаз, алты ай қыстағы табиғи ашық панасы» дегенге саяды. Қазақ байларының үйір-үйір жылқысы тоғайда жатады. Қазақ эпосында: «Тоғай толы жылқымды аш бөріге тапсырдым», – десе, «Шора батыр Нәрікұлы» хикаясында «Меңлісұлу ханға қарап бір хикаят бастады: – Әй, ханым, біздің ағамыз бай кісі еді, ағамыздың тоғыз тоғай жылқысы бар еді», – дейді.

Халық ақыны Кенен Әзірбай:

*Сапырған сары ожаумен сары қымыз,
Ысталған сары местен дәру қымыз.
Қарасаң, қалың тоғай толған жылқы,
Түн тұрмақ таба алмайсың қарап күндіз, –*

деген екен (Кенен Әзірбаев, Шығармалар жинағы).

Міне, тоғайдың дәріптеліп, жер-су атаулары кісі есімдері санатына еніп, Тоғай, Байтоғай, Тоғайбай делініп «паспортталуының» детерминациясы географиялық объектінің малшы өміріндегі алатын орнының аса маңыздылығы, ең алдымен жылқы шаруашылығымен байланысты екенінде. Осыған орай нәрестенің атын «тоғай толы жылқы айдаған бай болсын» деп Тоғайбай, Байтоғай, Тоғай деп ырымдап қоятын болған.

Қазіргі синхронды сипаттағы түсінігімізде тілдік санадағы Тоғай, Байтоғай, Тоғайбай есімдерінің аңғартатын меңзелімінің әуелде жылқы түлігімен байланысты қойылғаны ұмыт болып, жабық кодқа айналды.

Сол сияқты қазіргі кезде Мыңбай, Бүтінбай, Көпбай деген есімдердегі кодқа салынған ақпараттар үстірт ашылып, толып жатқан көптеген байлардың бірі сияқты ұғынылады. Алайда диахрондық тұрғыда этномәдени санада Бүтінбай, Мыңбай, Көпбай есімдерінің кодқа салынған ақпарат (мағына) меңзелімі ақ малға, яғни жылқы түлігімен байланысты болған. Этномәдени танымда мың қой айдаған, жүз түйе өрістеткен атақты дәулет иелерінен гөрі мың жылқылы байлардың қоғамдағы әлеуметтік мәртебесі әлдеқайда жоғарырақ тұрған. Өзге де түліктер сияқты жылқы малы қазақтың жесе, азығы; ішсе, сусыны; кесе, киімі болған дей отырып, қыл құйрықты алысты жақындататын көлігі, жорық пен жаугершіліктің малы деп, әскери стратегиялық тұрғыдан бағалаған. Осы жоғарыда айтылған қазақы есімдердің детерминанттары этностың аксиологиялық танымымен байланысты когнитивтік түзілімнің мәдени тілдік матрицасы болып саналады.

Дәстүрлі халықтық антропонимдердің мифологиялық таныммен байланысты тұстарында сан категориясының алатын орны ерекше екені байқалады. Санның ондық жүйесінде үш, бес, жеті тәрізді тақ сандар белсенді жұмсалған: Үшкөмпір, Үшбай, Беспай, Жетібай т.б. Жұп сандардың тасада қалып, тақ сандардың артықшылыққа ие болуын бұл тұста да эзотериялық наным-сеніммен байланысты екені аңғарылады. Байқауымызша, жорамалдауымызша, тақ сандарда ортаға (ішке) тартқыш тылсым күш, ал жұп сандарда сыртқа тепкіш тылсым күш басым болады. Алайда қолымызда, анығын айтқанда, нақты сүйенер дерегіміздің болмағанын мойындай отырып, бұл тұста егде жастағы коммуниканттардың сөздерінен байқап қалған ауызекі деректерге мән беріп, назарға алдық. Эзотериялық наным-сенім бойынша баласы тоқтамай, нәрестесі шетіней беретін әйел жаңа туған баласын қолына алып, емізбестен, ауыл-үйлердегі немере-шөберелі үш көмпірдің бұтының арасынан өткізіп, зияндасты адастыру ырымын жасатып, нәрестесінің атын Үшкөмпір деп қоятын болған.

Сондай-ақ егіз қозылар көбейсін деген ниетпен, баланың несібесі зор болсын деп, қой төлдеп жатқан кезде туған баланы қазақы дәстүрде Егізбай деп атаған. Баласы тоқтамай, шетіней беретін әйел ырым етіп, жаңа туған нәрестесін көпбалалы бабасал бәйбішенің етегіне салады. Ондай сәбидің аты жасалған ырым бойынша Етекбай атанады. Бұлар – әңгіме арасында айтылғандардан құлақ шалған деректер. Сондай әңгіме арасында бабасал бәйбішелердің «Үшбайдан кейін туған Беспайым ғой» дегендерінен жұп сан «төрттің» қолданылмай бұгулі қалғаны байқалады. Беспайдан кейін туғандар үшін жұп санның қаупі жоқ әрі қарайғы туған балалардың аты Алтыбай, Сегізбай немесе Көбейсін, Жүрсін деп «паспортала» береді.

Қазақтың антропонимдер жүйесінде гендерлік ұғыммен байланысты этномәдени дәстүр ерекше орын алғаны байқалады. Атап айтқанда, осы аталмыш түсініктің түйіні байырғы тілімізде ұшырасатын қатығымен туған деген сөздің мағынасымен шешіледі. Қазақтың бұрынғыдан қалған сөзінде үйелменді үйдің ұл-қызы аралас балаларын қатығымен туған екен дейді.

Ал әйел тек қыз туа берсе, бұрынғылар ұл тілейді, ұл туа берсе, қыз болса деп үміт етеді. Үйелменді үйдің ұл аңсаған тілеуін қыз балаларының Ұлтуған, Ұлбике, Жаңыл, Адас деген т.б. есімдерінен кодқа салынған мәдени семантикалық ақпаратты аңғаруға болады.

Ал көп ұлдан кейін жарық көрген нәрестесі қыз болса, Тәттігүл, Тәттібала, Оңлыбала тәрізді есімдермен атайды. Ұлдан кейін қыз туса деп дәмеленген ата-ана қызының атын Данагүл, Гүлдана, Ақерке деп қояды. Ұл туса, қонысымыз, іргеміз кеңейді деп Қонысбай, Іргебай қояды, қыз туса, өрісіміз ұзарады деп, Шолпан, Айкүміс, Айсұлу, Таңшолпан, Жұлдызай қояды. Бұндай есімдердің мәдени семантикалық детерминациясы тек эстетикалық сұлулыққа тән таным ғана емес, көшпенді елдің Ұлы Даладағы ұзын жолдың бағыт-бағдарын танытатын аспан шырақтары әлемімен де байланысты. Ұлын ұяға, қызын қияға қондыру ата-ананың парызы. Еркек кіндікті өлсе, сүйегі өз елінде, ұзын етек ұрғашы өлсе, сүйегі күйеу жұртында болады. Байырғы ата-бабалырымыздың «қыз – жатжұрттық» дейтіні сондықтан. Қыз жат болып кетеді деген сөз емес, сүйегі жатжұрттық деген этномәдени ақпарат кодқа салынған.

Қыз – тіршілігінде ешуақытта жат болмайды. Бегім Ана, Домалақ Ана, Найман Ана, Қызай Ана, Ақ Кете сынды аналар рулы елдің қарым-қатынасының ұйтқысы болып, Ана деген ең ардақ тұтар мәртебемен халықтың тарихи жадында сақталып қалды. Халықтың «астың дәмін келтірген тұз әулие, алысты жақындатқан қыз әулие» дейтін даналығын еске алсақ, Нағашыбай, Жиенбек, Жиенбай сынды есімдер еске түседі. Шамасы, мұндай этномаркерлік есімдер жаңа босанған қуанышты шақта төркін жұртынан келіп жатқан игі жақсы адамдарды немесе бірер балалы болып, төркіндеп барған аяғы ауыр әйелдің барған жерінде босанып нәрестелі болғанын меңзейтін мәдени семантикалық ақпарат десек дұрыс бола қоймас дей ойлаймыз. Бұл – үстірттегі түсінік. Ал этнотілдік сананың астырттағы ойтүзіліміне бойласақ, алыс-жақын жерлермен Еңсегей бойлы Есімхан талап еткендей, қыз алысып, қыз берісіп құда болысу, түптеп келгенде, этностың тілін, салт дәстүрін, дінін, дүниетанымын, рухани құндылықтарын бірегейлендіретін факторлардың бірі және бұл айтылған факторға С.Е.Малов айтқандай, қатынасқа қолайлы Ұлы Далаға тән географиялық фактор қосылады. Сондай-ақ алысты әп-сәтте жақындатқан жылқы түлігі де маңызды факторлардың бірінен саналады.

Ономастикалық корпустағы кейбір халықтық топонимдердің мәдени семантикасының берілуі. Халықтық топонимдер жүйесінің онтологиялық болмысы табиғат объектілерімен тікелей байланысты болғанмен, олар перцепциялық қабылдаудың сүзгісінен өтіп, этнотілдік ұжымның когнитивтік санасында қорытылып, қайта жаңғырған мәдени тілдік өнімдер болып табылады.

«Ақ» пен «қара» топонимдері. Халық тіліндегі ақ, қара, көк, қызыл, сары, қоңыр, жасыл тәрізді түр-түс атаулары, бұлардан туындаған сөздер түр-түс категориясы деп аталады. Ақиқат дүниедегі түр-түс әлемі этномәдени ұжымның сезімдік қабылдау сүзгісінен өтіп, ой-санада категория түрінде қорытылған мәдени-тілдік өнімдер санатына жатады.

Әуелде ақ, қара, көк, қызыл, *сары, қоңыр* сияқты т.б. түр-түс атауларын білдіретін сындық мағыналарының онтологиялық болмысы, түптеп келгенде,

табиғат объектілеріндегі әралуан спектрлі құбылыстармен байланысты болды. Тілдегі түр-түс табиғаттың өзінен еншіленді. Басқаша айтқанда, ақ түс қармен, қара түс түнмен, көк түс аспанмен, қызыл түс күннің арайлы шапағымен ассоциацияланды. Сөйтіп, этномәдени ұжымның тілдік санасында табиғат қойнауынан алған түр-түс атауының конвертерленген бейнесі (картинасы) қалыптасты.

Алайда этностың тілдік санасы табиғаттағы түр-түсті «тіркеп» қана қойған жоқ, оларды ойтанымдық (когнитивтік) өңдеуден өткізді. Айталық, күннің алқызыл арайлы шапағы, айлы түндегі аспан шырағының ақ сәулесі этнос санасында қайта жаңғырып, әсемдіктің, сұлулықтың нышанына баланып, Күн де, Ай да, Таңшолпан да сұлулықтың символына айналды.

Халықтық астрономикада Күнсұлу, Айсұлу, Таңшолпандардың кең тараған есімге айналуы сондықтан.

Сондай-ақ ақ көйлек, қара шапандағы «ақ», «қаралар» табиғаттағы түр-түске байлаулы болса, ақ ниет, қара ниеттегі «ақ», «қаралар» ойтанымдық (когнитивтік) әрекеттің барысында табиғаттың байлауынан босанып, екінші деңгейдегі өзгеше мағыналық құрылымға түскен атауларға айналды.

Қазақтың ежелден ел болып күл төккен ата жұрты, ата қонысы ұлы дария, өзен-сулармен байланысты қолдың бес саласындай бес өңірді алып жатады. Бес өңір деп отырғанымыз Ұлы даланың шығысы мен терістігіндегі Ұзын Ертіс, батысында Ақ Еділ мен Ақ Жайық, түстігінде Сыр Дария, түстік шығысында Жетісу.

Көңіл аударып, ден қоярлық бір ерекшелік осы аталған Ұлы сулар өңірінің барша жерінде түр-түспен аталатын халықтық гидроним, оронимдер соншалықты жиі кездеседі. Бұларды категория тұрғысында қарастыру дәстүрлі ономастикалық атаулардың халықтық географиямен байланысты танымдық қырларын аша түсуге мүмкіндік береді.

Халықтық гидронимде әйгілі Ертіс суы Ақ Ертіс және Қара Ертіс деп екіге бөлінеді. Оның Жайсаң (Зайсан) көлдің түстік жағалауына құйған жағы Қара Ертіс, Терістіктегі Бұқтырма жағынан ағып шығатын жағы Ақ Ертіс деп аталады. Бұндағы ақ атауының ертеректегі мағыналық құрылымында «ақ»- «ағу» деген қимыл, қозғалысқа тән семалық элемент болғаны байқалады. Ал Қара Ертіс атауындағы қара «үлкен» деген байырғы мағынаны білдіретін ахроматизм екені қара шаңырақ сөзіндегі «қарамен» салыстыру арқылы айқындалады. Сонымен қатар өзен-судың сол аймақтағы Сары Ертіс, Бала Ертіс деген түрлері Қара Ертістен олардың кішірек екенін аңғартса керек.

Сөйтіп, үш Ертістің атауы қазақ халқының ең байырғы фольклорлық мұрасы «Қозы Көрпеш-Баян сұлу» жырында айтылған.

«Ертістің басы – Қара дің» дей келіп, онда былай дейді:

*Алтайдың бауырында Сары өзен,
Сары Ертіс, Қара Ертіс бар айыр кезең.
Бала Ертіс және ағып қосылады,
Үстінен Зайсан көлдің кетер төмен.*

(ҚҚБС. Дәуіт нұсқасы)

Күнгей Алтайдың Қара Ертіске құятын түстік шығысында Қаба суының бір саласы Қара Қаба, екінші бір саласы Ақ Қаба деп аталады. Тау суының бойы қара тоғай, қалың жынысты болғандықтан Қара Қаба, ал Қабаға құйғанға дейінгі екінші саласы жазыққа таман жайыла аққандықтан Ақ Қаба аталған. Қаба гидрониміндегі қаба – «қалың», «ұйысқан» деген мағынаны білдіретін көне атау. Бұл атаудың этимологиясын жылқы түлігіне қарата айтылатын қаба жал; қаба құйрық деген фразеологизмнің «қалың» деген мағынасымен салыстыра қарауға болады. Қаба жеріндегі тағы бір гидроним Қаба Арасан деп аталады. Саз балшықты, минералға бай болғандықтан солай делінсе керек. Осы аталған байырғы топоним атауының Жетісу жеріндегі Қапал Арасанмен түбірі тектес объект екені байқалады. Қапа>қаба/қапа атаулары түбірлес сөздер. Қапал Арасан топониміндегі диахронды дыбыс өзгерістерді Қабалы Арасан>Қаба Арасан>Қапал Арасан түрінде көрсетуге болады.

Жайсаң көлдің Бұқтырмасынан қысыла шығып, кең жазыққа жазыла аққан ұзыннан-ұзақ тағы бір ерекшелігіне орай халық тілінде Ұзын Ертіс атанған.

«Ақ» атауымен келген гидронимдердің мазмұн межесіндегі ақпараттар халықтық географияға тән эмпирикалық білімді аңғартады. Орал тауының оңтүстігіндегі қырқалы, қыспарлы жерлерден түстікке қарай құлаған Жайық Жазыққа шығып, Ордың құйған сағасынан өте бере батысқа кілт бұрылып, Каспийге бет түзейді. Оған құятын үлкенді-кішілі өзен-сулар көбейіп, жайыла, жазыла аққан ұлы дарияның осы аралығы этнос санасындағы бейнесіне орай Ақ Жайық аталған деуге болады. Жайық, түптеп келгенде, жайылу, жазылу, жайылма сөздерімен түбірлес атау.

Ақ Еділ гидронимінің де географиялық ерекшелігі осыған ұқсас. Оның бастау алатын тұсы Жайықпен жақындау, біраз жер Жайықпен жарыса қатар түзеп ағады да, Жайықтан ертерек, сәл теріскейлеу батысқа кілт бұрылып кетеді. Ақ Еділге құятын өзен-су, бастау, бұлақтар көбейе келе дарияның су көзін байыта түседі.

Осылайша қос өзеннің этнос санасындағы ақиқат дүниемен байланысты тілдік бейнесі қалыптасқан. Бұндағы «ақ» та «ақ», «ағын», «ағыс» деген синкретті ұғымның атрибуты болып саналады.

Ұлы Абай мектебінің алдыңғы кезеңіне дейінгі аса көрнекті сыншыл ақындар санатындағы Дулат Бабатайұлы:

*Ақжайлауым Сандықтас,
Атам қонған қонысы.
Тұн асса тұтам шөп өскен,
Басылмайтын сонысы, –*

дейді.

Ақынның Ақжайлау деп қасиет тұтып отырғаны ағынды ақ бұлақ, ақ бастау, ақ қайнар, ақ тұмалары көп, берекелі, мал-жанға жайлы жердің тілдік бейнесі. Ақын айтып отырған бағалауыштық мағына – этномәдени ұжым санасындағы аксиологиялық танымның дала малшылығымен байланысты ерекше түрі.

Халықтық ономастикада ақ атауымен айтылатын топонимдердің мәдени семантикасының мазмұны ерекше екені ондағы кодқа салынған бүкпелі ақпараттарды ашқанда айқындала түседі.

Ақтоғай топониміндегі ақ «ағу», «ағын су» дегенді, ал тоғай осы судың бойы мен жағалауына біткен тал, терек, қайың, ырғай, шілік, бұтасын өскен аралас қалың ағаш дегенді меңзейтін танымдық ақпарат кодқа салынған. Аталмыш топонимнің мәдени семантикалық құрылымында кодқа салынған ақпарат бұл айтылғандармен шектелмейді. Ақтоғай сияқты тоғайлы жерлер жыл бойы қора көрмей, далада бағылатын ақ малдың, яғни жылқы түлігінің сан мыңдаған қосына қақаған аяз, қарлы борандарда ең қолайлы ықтасын болады, әсіресе аяқты малдың ығып кетуінен сақтайды. Бұндай жерлерді қазақ даласының кей өлкесінде Шұбар, Шұбарағаш, Шұбарқұдық деп те атайды. Ал малға пана бермейтін жарлауыт, жарқабақты өзен бойындағы тоғайлы жерлер халықтық ономастикада Жаманшұбар деп те аталады. Бұндай жерлердің Қаратоғай деп аталатынын байырғы мұралардан да кездестіруге болады.

Қаратоғай деп ішіне мал кіре алмайтын, қалың жынысты жерді атайды. «Жамандық қалың Қаратоғайдың ішіне Жақсылықты тастап кетті», – дейді ертегіде (Ә.Диваев, 34).

Ақсу, Көксу елді мекен, өзен-су атауы (гидронимдері) құрамындағы «ақ» пен «көк» компоненттері түр-түсті білдіретін кәдімгі ақ пен көк емес. Ақсудағы ақ та, Көксу дегендегі көк те қысы-жазы басынан қар кетпейтін биіктегі ақ қар-көк мұздың суы дегенді меңзейді. Атап айтқанда, ақ қар, көк мұздың суы деген тұрақты тіркес когнитивтік әрекетке тән үрдістің барысында ақ су, көк су болып қайта құрылымданған. Басқаша айтқанда, санадағы когнитивтік әрекеттің барысында форма мен мазмұн жағынан кірігіп (контаминацияланып), бүтін бір гидронимдік бірлікке айналған.

Ақсу, Көксу таулы өзеннің суы, суы тұщы болатыны, ағысы арынды келетіні, асып-тасып өткел бермей ағатыны, кей тұстағы өткелі тас кешу тайғақ болатыны халықтық гидронимнің мазмұн аясындағы танымдық ақпараттар болып табылады. Су тасқынынан кейін ақжоңқа ағаш қалдықтары, түбір, бұтақтар өзен жағалауларында көптеп кездеседі. Жергілікті тұрғындар оны салынды дейді. Өзен бойындағы тал-теректерге балта тигізбей, салындыны жиып-теріп, отын-суға жаратады. Міне, бұндай азды-көпті ақпараттар өзен бойындағы жергілікті тұрғындар мен табиғат арасындағы қатынастың, байырғы кезде, техникалық прогресс заманына дейін, қандай сипатта болғанын аңғартады.

Ақсу, Көксу өзендерінен бірдиірмендік суы бар тоған қазылып, су диірмендері істеп тұрды. Диірменнен шыққан су айдалаға кетпес үшін егістік жерлерге тартылған тоғанмен суармалы дақылдар өсіруге пайдаланылды. Болмаса, едәуір жерге дейін қазылған тоғанмен өзен арнасына қайта қосылатын су жүйелері жасалады.

Сөйтіп, өзен алабының ежелгі тұрғындарының табиғатпен қатынасында су көздерін ысырапсыз пайдалану ұстанымының ғылыми-техникалық революцияға дейін қандай болғанын судың да сұрауы бар; судың да сұрауы зор деген мақалдан (паремиологизмнен), сонымен қатар Су патшасы – Сүлеймен:

Су иесі – Сүлеймен дейтін мифологияға тән иррационалдық таныммен байланысты түсініктерден де айқынырақ аңғаруға болады.

Ежелгі қазақтардың ата жұртында ақ сөзімен келетін гидронимдердің тағы бір ерекше түрі бар. Қазақтың бағзыдан бауыр басқан өлкелеріндегі Ақкөл деп аталатын халықтық гидронимдері этнотілдік ұжымның ерекше географиялық біліміне тән рационалдық сипаттағы атаулар. Атап айтқанда, өзен-сулардың бір жағалауынан көлге құйып, екінші жағалауынан ағып шығып жататын ерекше географиялық гидрообъекті халықтық білім өніміне тән атауға жатады. Қазақтың ежелгі Талас өңіріндегі Ақкөл, байырғы Торғай өлкесіндегі Ақкөл гидронимдеріндегі ақ сөзінің түр-түске қатысы жоқ, бұндағы ақ сөзі – әрі зат, әрі етістік мағынасындағы байырғы кезеңдегі синкретті түбір. Оқулықтарда айтылғандай, көш – әрі зат, әрі етістік, ақ та солай әрі зат, әрі етістік.

Ал Ақбұлақ, Аққайнар, Ақтұма тәрізді су көздері атауларындағы ақ объектінің «ағу», «ақ»-пен байланысты сипатын аңғартады. Гидроним атауының құрамындағы ақ синкретизмінің мазмұнында «қозғалыс», «жылжу» деген мағыналық компонент Аққұм топонимінде де кездеседі.

Құмның бұндай түрінде оны ұстап, бекітіп тұратын қара отқа жататын өсімдіктер тым сирек кездеседі де, құмы сусымалы, сырғымалы болады.

Қаратау дегендегі қара қар ұзақ жатпайтын географиялық ерекшелігімен байланысты. Мысалы, жер қара, күн жылыда деген тұрақты сөз орамындағы қара «қарсыз» ұғымының референті. Бұндағы қара-ның кіндігі жерге байланып, өзінің табиғи түр-түсті білдіретін төлтума мағынасынан алшақтап кеткен. Қаратау оронимі түркі халықтарының ел болып бауыр басқан мекендерінде жиі кездеседі. Қазақстанның түстігінде Ақ Еділ-Жайық қос өзені аралығында, Қырым, Балқан жерінде де ұшырасатын ороним. Қара – жалпытүркілік ойтанымда жердің тұрақты атрибутикалық белгісіне айналған ерекше феномен.

Байырғы қазақ елінің дәстүрлі малшылық кәсібінде қара сөзімен келетін қыстаулар жиі кездеседі. Мысалы, Желдіқара атауындағы қара – таудың қарайып, қар аса көп жатпайтын күнгей беткейі. Ал желді сөзі – түстіктен терістікке қарай соғатын жел. Бұндай жел жердегі қарды үрлеп, қағыл қар етіп, жұқартып жібереді. Қара малдың – қой-ешкі, түйе, сиырдың жайылымы қыстау маңындағы қара отты жерде болады, бірақ сиыр малының өрісі өленді жерде болатындықтан бөлектеу бағылады.

Міне, бұдан байқайтынымыз, күйіс малының қара мал аталуы олардың жайылымы қара от болғандықтан деуге болады. Оты басқа болғанымен, жу-сауы бір, әрі түйе, қой-ешкілер сияқты күйіс қайыратын болғандықтан, сиыр да қара мал санатына жатады. Қара мал, жылқы сияқты емес, су таңдамайды, көлкіп жатқан қара су, іркінді суларға бас қойып, іше береді. Қара малдың қара аталуы қара от дегендегі қарамен, ал қара оттағы қара атауы жер ұғымымен кіндіктес. Философия терминімен айтқанда, қара атауы қара малдың да, мал отының да, мал оты өсетін жердің де атрибутына айналған. Түптеп келгенде, бұл қаралардың бәрінің санадағы логикалық қисыны – жер дегенге саяды. Әдетте қара малдың жайылымы қырқа, қыратты, таулы жерлердің күнгей беткейінде болады. Қары желмен қағыл болып, қара оттың басы қар бетінен шығып жатады. Жайылымның кейбір жерлері қараң-құраң тартып тұрады. Қара малдың

бұндай қыстаулары Жетіқара, Үшқара, Желдіқара сияқты дәстүрлі атаулармен аталады.

Қарашоқы дегендегі қара түр-түс емес. Олай дейтініміз – халықтық метеорологияда қара суық дейтін терминдік ұғым бар. Бұндағы қара «жер»-мен байланысты ұғым, яғни «қарсыз суық, жерге қар түспеген кездегі суық».

Қарашоқы атауындағы қара географиялық объектінің түстік бетті ала, қары аса ұзақ жатпай, еріп, жер қарайып кететін тау алабының ерекшелігін аңғартады. Қара атауы тілдік санада жердің тұрақты атрибуты, атап айтқанда, «заттың» орнына оның атрибутының жұмсалуды. Ежелгі қазақ жерінде Қарасу, Қарабұлақ гидронимдеріндегі қара су көзінің жерден шығып жатқанын аңғартады, яғни заттың орнына оның «меншікті» атрибутының қолдануы тілдік санадағы когнитивтік әрекеттің ерекше өнімдері.

Қарасу, Қарабұлақ гидронимдеріне тән тағы бір өзгешелігі ағысы білінбейді немесе тым баяу болады. Бұндай өзендердің ағысы су астында болып, түбіндегі шөптесін, балдырлардан байқалып тұрады.

Ақмола, Ақмешіт топонимдері құрамындағы ақ атауының мәдени семантикалық мазмұнында кодқа салынған ақпарат халықтың діни білім, танымымен байланысты деуге болады. Агиотопонимдегі ақ атауы «әулие», «ақсүйек» ұғымының референті болып саналады. Бұндай қасиетті деп танылған агиотопонимдердің мәдени семантикасында дәстүрмен байланысты ақпарат бар екені байқалады. Әулие, ақсүйектердің басына тұрғызылған қорған, күмбез әдетте күре жолдың бойына таяу биіктеу жерде болады. Жүргіншілер құран бағыштап, «қияметтің күні жақын» деп аруақты тыныштандырып, бет сипап өтеді.

Сондай-ақ агиотопонимдердің қатарында Қарамола деген атау да кездеседі. Бұндағы қара, шамалауымызша, «қарасүйек» ұғымымен байланысты, қара халықтан шыққан тұлға деген танымдық ақпарат кодқа салынған. Қарамола да күре жолдың бойында болады. Дәстүр бойынша жолаушы, жүргіншілер жүрістерін баяулатып, бет сипап, «аруақты тыныштандырып» өтеді.

Ақмола, Қарамола сияқты объектілер ұзын жолдағы қазақ үшін жол шама-сын межелейтін әрі адастырмас ерекше белгі болып табылады. Бұндай ескерткіштерді жол үстіндегілердің «Темірқазыққа» балайтынын классик қаламгер Ғ.Мүсіреповтың шығармаларынан да ұшыратамыз.

Бұл орайда ұлы Абайдың:

*Жабырқаңқы көрінер жазған сөзім,
Жұрт жабылып тоқтатар оған көзін.
Жолаушы жол үстінде тамаша еткен,
Секілді өткен жанның бір күмбезін, –*

деген өлең жолдарынан да дәстүрлі наным-сенімге тән ақпаратты аңғарғандай боламыз.

Сонымен, біз аталмыш зерттеуімізде ақ, қара атауларымен келген топонимдердің мәдени семантикалық ерекшеліктеріне талдау жасадық. Сондағы байқалатыны «жер», «өсімдік», «мал» ұғымдарының арасындағы ішкі мағы-

налық байланыстың логикалық қисынға құрылуы. Олай дейтініміз: қазақ тілінің байырғы лексикалық қорындағы қара мал (түйе, қой-ешкі), ірі қара (түйе, сиыр); ақ мал (жылқы) атаулары түр-түсті білдірмейді, атап айтқанда, қара малдың жайылымы, оттайтын оты – қара от, сондықтан қара мал атанған. Қара отқа қара барқын, қара соран, жантақ, күйреуік, сексеуіл, жыңғыл, тораңғы, кекіре, алабота, көкпек, теріскен, дүзген жатады. Ал ақ малдың (жылқының) жайылым оты ақ от, сондықтан жылқы түлігі ақ мал атанған. Ақ отқа, яғни ақ малдың отына бидайық, ажырық, көде, бетеге, тарлау т.б. жатады.

Ал оттың қара от аталуы қырат-қырқалардың күнгей бетінің қарайып жататын ерекшелігімен байланысты болса, оттың ақ от аталуы теріскей беттегі жерлердің оты бозғылт тартып жататын ерекшелігімен байланысты. Бұндай жерлердің шөбін боз от, боздауық деп те атайды. Бұдан малшы қауымның халықтық географияға тән рационалдық танымына тән білімін айқын аңғаруға болады. Ақ, қара, қызыл, сары, қоңыр, шұбар... т.б. түр-түс атауларының мағынасы дәл осы күйде – табиғат қойнауынан сезімдік қабылдау арқылы алынған күйде – қалып қоймайды. «Түпнұсқалық» мағына мүлде өзгеріп, когнитивтік сананың екінші деңгейіндегі тілдік құрылымға айналады. Бұлар фразеологиялық мағына, символдық мағына, эталондық мағына деп аталады. Мысалы сары аяз, сары жел, сары желіс, сары уайым дегендегі «сарылар» менталдық әрекеттің (актінің) барысында табиғат аясындағы байлаудан босанып, когнитивтік деңгейдегі мәдени семантикалық құрылымға айналған. Бұндағы мәдени деген анықтауыштық термин субъектінің (этностың) табиғаттық емес, өзіндік қолтума «өнімі» дегенді білдіреді. Олай дейтініміз, аяз, жел, уайым, желіс дегендерде түр-түс болмайтыны белгілі. Міне, бұл құбылыс бір жағынан бір кездердегі этномәдени ұжым санасындағы танымның эволюциялық даму үрдісі болса, екінші жағынан тілдің лексика-фразеологиялық жүйесіндегі құрылымдардың да эволюциялық дамуы болып табылады.

Тілдік санадағы когнитивтік әрекеттің барысында сары атауымен келген сары аяз, сары жел, сары уайым т.б. фразеологиялық бірліктердің құрамындағы сары «аса ұзақ», «көпке дейін созылған» дейтін мәнділікті аңғартатын ақпарат когнитивтік әрекеттің барысында пайда болған. Сары жел «бірнеше күн», сары аяз – «ұзаққа созылған аяз», сары уайым – «ұзақ уақытқа созылған уайым» деген фразеологиялық мағыналарды білдереді. Ал сары атауына тән бұл ақпарат өзінен-өзі пайда болған жоқ. Мысалы, сарымай дегендегі сары түр-түсті, сонымен қатар ұзаққа сақталатын дегенді де білдіреді. Қарынға салынған сарымай ұзаққа сақталады; сондай-ақ сары тоқым дегендегі сары тоқымның ұзақ уақыт ұстала сарғайып кеткенін байқатады. Сөйтіп сарының мағыналық құрылымында пайда болған «ұзақ», «ұзақ уақыт» компоненті когнитивтік санада актуалданып, түр-түс компонентінен гөрі басым сипат алып кетеді. Түр-түс компоненті когнитивтік санада шеткері аймаққа ығысады.

Топонимдер құрамында кездесетін ақ, қара сияқты сары, қоңыр түр-түстерінің перцептивті қабылдауда өзінің табиғи болмысына сәйкес келмей ауытқулары ой-санадағы когнитивтік әрекеттің екінші деңгейдегі, яғни менталды әрекеттің мәдени танымдық өнімдеріне жатады.

Құрамында сары, қоңыр атаулары бар топонимдердің өз ерекшеліктері бар. Сары, қоңырлардың табиғи болмысынан қол үзіп, ахроматизмге айналуының когнитивтік механизмін салыстыра талдау арқылы айқындауға болады.

Айталық, қазақтың ас-тағам дайындау дәстүрінде сиыр сүтінен айырған майды ұзаққа сақтау үшін оны малдың ерекше өңделіп, кептірген қарнына, бүйеніне салып, аузын тігіп қояды. Қарындағы сарымай өзінің табиғи түр-түсін әрі оған қоса ұзақ сақталған деген ақпаратты аңғартады.

Қазанда қайнатқан іркітті құрт алу үшін оны дорбаға салып, сары суын ағызады, ақыр соңына дейін сары су тамшылап ұзақты күн ағады. Міне, осы көріністер перцептивті қабылдауда «ұзақ», «ұзаққа созылған» деген мағыналық компонент пен сары аяз, сары уайым, сар(ы) желіс, сары таң, сары бел, сары ауру дегендегі сарылар өзара төркіндес. Басқаша айтқанда, аталмыш идиомалардағы сарылар когнитивтік өңдеуден өтіп, «ұзақ», «ұзаққа созылған» деген мәдени семантикалық құрылымға айналған. Бұндағы мәдени семантика деген анықтауыштық термин субъектінің (этностың) табиғаттық емес, өзіндік қолтума «өнімі» дегенді білдіреді, себебі аяз, жел, уайым, желіс дегендерде түр-түс болмайтыны, екінші тілге сөзбе-сөз аударылмайтыны бұлардың идиомалану қасиетімен тікелей байланысты.

Сол сияқты Сарысу, Сарыбұлақ, Сарыжайлау, Сарыарқа топонимдеріндегі сарылар түр-түсті білдіретін табиғи мағыналарынан ажырау (бифуркациялану) арқылы «ұзақ», «ұзаққа созылған» деген қолтума мағынаға ие болғанын ойтанымдық әрекеттің идиомаланған өнімдері деуге болады. Бұл – этномәдени тілдік ұжымның табиғатпен алыс-берісі, табиғаттан алғанын қайта өңдеп, мәдени өнімге айналдыруы.

Айталық, халықтық топонимдердің құрамында кездесетін сары компоненті ұзаққа созыла, үздік-создық болып ағатын, кей тұстарының суы тартылып, курап, жылға, жауын-шашын суларымен ара-тұра қайта байитын өзен-сулар тобына қатысты ерекшеліктерді білдіреді. Ал бұндай сулардың бойындағы тоғай гидронимнің осы өзгешелігіне орай Сарытоғай деп аталады. Сарысу, Сары өзендердің жазық жерлердегі бойында иірім, іркінді сулар жиі кездеседі. Ондай Сарысу бойында өскен қоға, қамысты жерлерді Сарқамыс деп атайды.

Сарыжазық атауының тілдік матрицасында әртүрлі ұғымға қатысты ақпараттар құрылымданған. Сарыжазық тау етегіне таманғы кең, ұзыннан-ұзақ жазықтықтағы жайлаудың атауы. Жайлаудың бұл түрінде кедей немесе әл-ауқаты орташалау ауылдар отыратыны қауымның әлеуметтік сипатын аңғартатын бүкпелі ақпаратқа жатады.

Жетісулық қадірменді айтыскер халық ақыны Қалқа Жапсарбайұлының өлеңінде:

*Жайлауы біздің елдің Сарыжазық,
Көпшілік кедейлерге болған азық,-*

деген жолдар бар. Бұндай жерлердің жайылым оты тұяккесті болып ескіре бастағаннан-ақ ауыл ел күзеуге көше бастайды. Ал Төржайлауда, одан да биік Кержайлауда бай ауылдар отырады, бұндай топологиялық кеңістіктердің шөбі

шүйгін, бұлақ, бастаулары таза, түні салқын, мал мен жанға қолайлы болып келеді.

Топонимдердегі сан атаулары. Ұлы даланың жер бедеріндегі санмен айтылатын төбелі, қырка-қыратты жерлері (оронимдер), өзен-су алаптары (гидронимдер), олардағы нәбататтар (фитонимдер) мен хайуанаттар (зоонимдер) дүниесі халықтық топонимдердің тілдік бейнесінде ерекше орын алады.

Атап айтқанда, Үштөбе, Үшқоңыр, Үшқағыл, Үшқара, Бестөбе, Жетіқара тау жоталары, төбе, қырка-қыраттарының; Үшбұлақ, Бестамақ, Жетісу, Жетісай өзен-су, көлдерінің; Бесағаш, Үштерек, Бестерек т.б. өсімдіктер дүниесінің ерекшеліктері мен сан атаулары компоненттерінен құралған топонимдер де сезімдік қабылдаудың сүзгісінен өтіп, ой-санада құрылымданған матрицасы мәдени семантикалық өнімдерге жатады.

Аталмыш топонимдердің тілдік матрицасында кодқа салынған әртүрлі ақпараттарға тән ерекшеліктердің бар екені байқалады. Үштөбе, Бестөбе топонимдеріндегі үш, бес объектінің сезімдік (перцептивті) қабылдауда өзге төбелерден оқшауырақ көзге шалынатыны туралы ақпарат кодқа салынған, яғни ұғымның тілдік матрицасы қарапайым ақпаратты аңғартады.

Ал Бесағаш, Үштерек, Бестерек дегендердегі «ағаш, терек өзен-сулар бойында әр тұста шоғырланып көзге шалынатын ағашпен» байланысты ұғымдық ақпарат, яғни ағаш, терек атауының тасасында «өзен-су» тұрады. Басқаша айтқанда, Бесағаш, Бестерек, Үштерек атауы – фитонимдер мен гидронимдердің санада контаминацияланған күрделі ұғымдардың тілдік матрицасы.

Сондай-ақ Үшқоңыр, Жетіқоңыр дегендегі үш=, жеті= сандары – сезімдік қабылдаумен сәйкес, яғни болмыс-бітімі айрықша көзге шалынып тұратын тау жоталары мен төбелер тізбегі. Ал бұндағы =қоңыр атауы салқын-самалды, мал-жанға жайлы деген ақпаратты аңғартады. Оны қоңыр салқын деген фразеологизммен салыстыру арқылы айқындай түсеміз. Сонымен, Үшқоңыр, Жетіқоңыр атаулары «таулы, төбелі, қырка-қыратты жерлерге тән, жаздың ыстық айларында малшы қауымның жайлаулық жері» деген ұғымдық мазмұнның күрделі тілдік матрицасы болып саналады.

Осылар тәрізді Үшқара, Жетіқара атауларындағы үш, жеті сандары таулы, төбелі, қыратты, қырқалы жерлердің санға сәйкес көзге шалынатын биіктеу тұстары, ал қара төбе, таудың қарайыңқы күнге беткейі деген ұғымдық мазмұнның күрделі тілдік матрицасы деуге болады.

Сан атауларымен айтылатын топонимдердің тағы бір атап айтарлық ерекшеліктері олардың тек қана, бірден басқа, үш, бес, жеті, тоғыз сияқты тақ сандармен қолданылатыны. Екі жұп саны тек қана Екібастұз атауында ғана кездеседі.

Сондай-ақ бір сөзі халықтық топонимикада жаңғыз/жалғыз түрінде айтылады: Жаңғыз төбе, Жаңғыз ағаш т.б. Осы сияқты екінші орнына қос, телі, егіз тәрізді жұптық ұғымды білдіретін атаулар қолданылады: Қосшалқар, Телікөл, Қособа, Егізкөл т.б.

Халықтық топонимдерде тақ сандарға артықшылық беріп, жұп сандардан «қашудың» детерминациясы ежелгі эзотериялық наным-сеніммен, яғни табиғатта жоқ апперцепциялық құпия күштерге сенумен байланысты деуге болады.

Үш санымен айтылатын кейбір топонимдердің ұғымдық мазмұнында объектінің үшемдік жүйеге тән мәдени семантикалық ақпараттардың кодқа салынғаны байқалады. Кодқа салынған бұндай бүкпелі ақпараттарды ашатын болсақ, мысалы, қатарласа жатқан Үшқағыл жоталарының Үлкен Қағыл, Орта Қағыл, Кіші Қағыл деп аталатынын, сондай-ақ Үшалматы топонимінің мазмұндық құрылымы Үлкен Су, Орта Су, Кіші Су деп аталатын, жоғарыдан төменге жарыса, қатарласа аққан үш өзеннен тұратынын аңғарамыз.

Тіл – өзінің онтологиялық болмысы жағынан табиғаттың қайталамасы. Табиғаттағы заттар мен құбылыстар, кеңістік пен уақыт, бағыт-бағдар, қарама-қарсылық, тыныштық пен қозғалыс, түр мен түс, сан-мөлшер, ыстық-суық, қатты-жұмсақ, дыбыс пен дауыс т.б. категориялар – тілдің өзімен бірге жасасып келе жатқан табиғи болмысы.

Табиғат объектілері сезімдік қабылдаудың сүзгісінен өтіп, танымдық этикеттің барысында атауларға ие болған. Айды «ай», Күнді «күн», барысты «барыс» атандырып, тілдік таңбаға айналдыруы адам санасындағы ойтанымдық, яғни когнитивтік әрекеттің бірінші деңгейі. Ал, жай ғана қарапайым мысалмен айтқанда, адамның Күнді «сұлулық», Айды «сұлулық», барысты «күштілік» деп тануы когнитивтік әрекеттің (актінің) екінші деңгейдегі, яғни табиғат дүниеліктерін мәдени тілдік өнімге айналдыруы болып табылады. Ономастикалық, оның қатарында антропонимдік атаулар ойтанымдық әрекеттің екінші деңгейіндегі түзілімінде өңдеуден өткен тілдік өнім тұрғысынан қарастырылды.

Сонымен, зерттеуде халықтық антропонимдердің, топонимдердің мәдени семантикалық детерминациясы ұлттың дүниәуи (материалдық) және рухани дүниеліктерінің базасы негізінде ашылды.

Елдің дәстүрлі мал шаруашылығы, кәсібі, табиғат объектілері, тұрмыс тіршілігімен байланысты материалдық құндылықтар жөніндегі рационалдық таным-білімдері антропонимдік атаулардың детерминациясы болды.

Ұлттың рухани дүниесіндегі этикалық таным-білімдермен байланысты адами әмбебап құндылықтар аясында қалыптасқан антропонимдер ономастика саласында ерекше орын алады. Бұндай атаулардың негізгі детерминациясы этикалық құндылықтар деп танылды.

Халықтың ономастикалық қорындағы антропонимдер, топонимдер ұлттың материалдық, рухани дүниесіндегі таным-білімге айналған құндылықтармен байланысты болып, этномәдени сананың кеңістігінде мағына, ұғым, концепт түріндегі менталды (ойтанымдық) түзілімдер форматында орныққан. Ал осы аталған ойтанымдық түзілімдердің тілдік матрицалары болып саналатын атау-сөз, фразеологиялық бірліктер, паремеологизмдер (мақал-мәтелдер), нақلياتтар халықтық антропонимдер жүйесінің қалыптасуының базасы болған.

3.2 Ономастикалық ішкорпус жасаудың әлемдік тәжірибесі

Қазіргі ғылым мен білім берудің т.б. қоғамдық өмірдің барлық салаларында IT технологиялардың қолданылуы күн озған сайын жеделдеп дамуда. Осыған орай заманауи технологиялардың жетістіктерін ғылыми зерттеулерде

пайдалану және «Цифрлы Қазақстан» бағдарламасының талап қажеттіліктері аясында қарастырып отырған қазақ тілін цифрлау мәселесі өте өзекті болып отыр. Корпустық лингвистика – мәтіндік базаны түзумен, оны әзірлеумен және қолданумен айналысатын белсенді дамып келе жатқан отандық тіл білімінің жаңа бағыты. Қазіргі уақытта әртүрлі тілдерде көптеген корпустар бар, оларды қалыптастыру, белгілеу және статистикалық мәліметтерді шығару технологиясы жетілдіріліп келеді. Жалпы корпустық лингвистиканың негізгі өзегі әрі зерттеу объектісінің бірі ұлттық корпус болғандықтан, әр тілдегі корпус әзірлеу мәселелері өзекті болып отыр. Қазақ тілінің ұлттық корпусы да басқа тілдегі корпустармен салыстыра қарағанда өзіндік жолмен біршама дамып, лингвистикалық белгіленімдермен толығып келеді. ҚТҰК-ға жалқы есімдердің де базасы еніп, ономастикалық ішкорпус қажет белгіленімдермен толығытыны анық. Ал корпустық ономастиканың қазіргі міндеті – әртүрлі жанрдағы мәтіндердің құрамындағы онимдерге белгіленім енгізу. Қазіргі таңда Қазақстан Республикасының ұлттық картографиялық-геодезиялық қоры жүйелеген 168000 (жүз алпыс сегіз мың) топонимдік атауды қамтитын microsoft access форматында сақталған электрондық база әзірленген [73]. Аталған база әр 6 ай сайын жаңа атаулармен жаңарып, ескі атаулары шегеріліп отырады. Алайда бұл база интернетте жүйеленбегендіктен, көпшілікке қолжетімсіз болып отыр. Корпустық ономастиканы дамыту міндеті тек топонимдерді ғана емес, сондай-ақ онимдердің басқа да түрлерін қамтитын, Қазақстанның ономастикалық қоры туралы жинақталған ақпаратты тек синхронды ғана емес, ретроспективті түрде де ұсыну үшін көп функциялы жалқы есімдер корпусын құруды көздейді.

Тіл білімінің барлық саласы компьютерлік құралдармен өңделіп жатқан тұста ономастикалық ішкорпусты да құрастырудың маңызы ерекше. ҚТҰК-да (qazcorpus.kz) тіл бірліктерінің барлығы көрініс тауып, арнайы шартты таңба беріліп, аннотацияланған. Тіл бірліктеріне (сөз, сөз форма, сөз тіркесі, сөйлем, мәтін) морфологиялық, фонетикалық, синтаксистік, семантикалық белгіленімдер жасалған. Алдағы уақытта ономастикалық атауларды да арнайы шартты белгілермен сипаттап беру – Ұлттық корпусты кемелдендіріп қана қоймай, ономастиканың дамуына да серпін береді. Ономастикалық ішкорпус – электронды түрдегі ономастикалық мәліметтер қамтылған ақпараттық-анықтамалық жүйе. Онда Қазақстанның топонимдері туралы лингвистикалық ақпаратпен қамтамасыз етілген мәтіндер жинағы қамтылған. Ономастикалық корпус Ұлттық корпустың бір бөлігі болғандықтан, оның пайдасы мен үлесі өзара сәйкесімді. Мәселен, корпустар арқылы тілдің дамуын бақылауға болады. Сол сияқты жер-су атауларының өзгерісін, мәтіндегі қолданысы мен кездесу жиілігін айқындауға болады. Сондай-ақ, тілдегі жалқы есімдердің өзгерісі тіркеліп, олардың кезең-кезеңге тиесілі қоры қалыптасады.

Зерттеу жұмысын жүргізу алдын әлемдік корпустарға шолу жасалды, олардағы ономастикалық атаулардың қалай аннотацияланғанына назар аударылды. Қазақстан Республикасының 2 облысы, атап айтқанда, Ақмола және Алматы облысының топонимдері ономастикалық корпустың бастамасы ретінде жинақталды. Оларды сипаттау, талдау жұмыстары жасалды және де алынған материалдар ономастикалық ішкорпус құруға негіз болды. Материал-

дар Қазақстан Республикасының географиялық атауларының мемлекеттік каталогынан [74] алынды. Біз құрастырып отырған ономастикалық ішкорпус – Ұлттық корпустың тәуелсіз бір бөлігі. Сондықтан да біз Ұлттық корпустың тәжірибесі мен ұстанымдарына сүйене отырып, ономастикалық ішкорпусты құрастырудамыз. Мұнда топонимдердің этимологиясы, семантикасы, грамматикасы, географиялық орналасуы жөнінде ақпараттар жинақталды. Осы тектес ақпараттар Т.Жанұзақ, А.Әбдірахманов, Б.Бияров сынды ономаст ғалымдар мен топонимика ғылымына қатысты энциклопедия, сөздіктер, географиялық анықтамалықтардан алынды. Мәселен, әрбір топонимнің атау ретінде қойылу себебі, мағынасы, этимологиясы мәдени семантика бөлігіне кіріп, осы топтың өзегін құрайды. Бұдан бөлек әрбір атауға қатысты тарихи, әдеби және мәдени-экономикалық т.б. қосалқы мәліметтер қамтылады.

Ана тіліміздегі сөздіктердің, энциклопедиялардың т.б. мәліметтердің көп болуы, әрі олардың электронды түрде болуы корпусты құрастыруға үлкен септігін тигізеді. Өйткені корпус дегеннің өзі мәліметтер қоры. Тіл білімінің жаңа салаларының бірі, корпустық лингвистика осындай мәтіндік корпустарды әзірлеумен, құрумен және пайдаланумен айналысады. А.Қ.Жұбанов және А.Ә.Жаңабекова осы саладағы ғалымдардың пікірін жинақтай келе, тілдік корпусқа төмендегідей анықтама береді. «Корпус дегеніміз – әр тілдегі электронды пішінге келтірілген, яғни бір басқару орталығынан автоматты түрде жұмыс істейтін, лингвистикалық ақпараттарды қамтитын мәтіндер жинағы» [12, 13]. Э.Д.Сүлейменова да осыған ұқсас түсініктеме береді. «Корпус – бұл мәтіндерді электронды түрде жинауға негізделген ақпараттық-анықтамалық жүйе. Анықтамадан айқындалып тұрғандай, біріншіден, корпус – бұл мәтіндерді электронды (цифрлендірілген) түрде жинау. Екіншіден, корпус анықтамалық-ақпараттық жүйе ретінде қолданыла алады.

Цифрлендірілген мәтіндердің базасы корпус белгісінің көмегімен арнайы өңделеді. Дәл осы белгі корпусты түбегейлі тілдің қолданылуы мен өзгеруі жайлы ақпараттың жаңа және объективті дереккөзіне айналдырады. Белгі – мәтіндердің және олардың бірліктеріне тән қасиеттер жайлы арнайы лингвистикалық ақпарат беретін корпустың негізгі сипаттамасы. Корпустың белгісі дәлірек және әрқилырақ болған сайын оның әлеуметтік, ғылыми және оқыту құндылығы жоғарырақ болады» [75, 371]. Ал *The Oxford English Dictionary* сөздігі корпусқа «Лингвистикалық талдауға негізделген жазбаша немесе ауызша материалдардың жиынтығы» деген анықтама береді [76].

Әлемдік лингвистикадағы ұлттық корпустардың ерекшеліктеріне қысқаша шолу жасаймыз. Ең жетілдірілген, әлеуеті жоғары әлемдік танымал корпустар қатарына Британ корпусы, чех тілінің корпусы жатады, орыс тілінің корпусы да едәуір жетілген деуге болады. Төменде ірі көлемді ұлттық корпустармен қатар әр алуан тілдердегі арнайы қызметтегі функционалды корпус түрлері бар. Корпустық лингвистика негізінен мәтіндік корпус жиынтығы болса, әлемде одан өзге алуан түрлі мультимедиялық (сериалдар, фильмдер, комикстер, мультфильмдер, түрлі сипаттағы телебағдарламалардың видеожазбалары және т.б. атай беруге болады) корпусты қамтиды. Мәселен, Колорадо университеті кітапханасының көрсетуінше, тілдік, мәтіндік корпустардың

ішінде Brown University Corpus of American English корпусы компьютермен қолданысқа арналған алғашқы цифрлендірілген корпус. Браун корпусы 1964 жылдан басталып жасалған.

Ал британдық ұлттық корпус (British National Corpus-BNC) – XX ғасырдың аяғынан бастап британдық ағылшын тілінің кең спектрін ауызша да, жазбаша да көрсетуге арналған көптеген көздерден алынған жазбаша және ауызша сөйлеу үлгілерінің 100 миллион сөздің жиынтығы.

Ағылшын тілді академиялық ортада қолданылатын өзге де көптеген іш-корпус түрлері бар, мәселен, BYU Corpora – Brigham Young University Corpora. CU – BYU Corpora тіркелген кітапхана болып табылады. Айрықша нұсқаулары мыналар: толық мазмұнға қол жеткізу үшін жүйеге кіру қажет. Тегін жазба үшін colorado.edu электрондық пошта мекенжайын қолдану арқылы тіркелуге болады. Деректер жиынтығы жүктеу үшін қолжетімді емес. Бұл дерекқорға кампустан тыс кіру үшін VPN пайдалану қажет. Корпоға-ға мыналар кіреді: ғаламтордағы жаңалықтар (NOW), ғаламдық веб-ағылшыны (GloWbE), Уикипедия корпусы (Wikipedia), Хансард корпусы (Hansard), интернеттегі ерте ағылшын тілі кітаптары (Early English Books Online), қазіргі америкалық ағылшын тілі корпусы (Coca), америкалық ағылшын тілінің тарихы корпусы (COHA), АҚШ Жоғарғы Сотының пікірлер корпусы (Corpus of US Supreme Court Opinions), TIME журналының корпусы, америкалық сериалдар корпусы, Британдық ұлттық корпус (BYU-BNC), Страти корпусы (Канада), CORE корпусы, Португал тілінің корпусы.

CHILDES – (Child Language Data Exchange System) корпусы – бұл балалардың бірінші тілді меңгеруіне бағытталған дыбыстық, бейне және мәтіндік файлдарды қамтитын мәліметтер базасының жиынтығы.

Келесі қытай-ағылшын параллельді корпустарының Chinese-English Parallel Corpora жүктемелерін TranslateX зерттеушілері мен тілшілері көпшілікке пайдалану үшін жасаған. Корпустар Гонконгтың қаржы және заң салаларын қамтитын сапалы екі тілді мәтіндерден тұратын сөйлемдердің жұптарынан тұрады. Дереккөздерге үкіметтік заңдар мен нормативтік актілер, қор биржасының хабарландырулары, қаржылық ұсыныстар туралы құжаттар, нормативтік құжаттар, нормативтік нұсқаулықтар, корпоративтік құрылтай құжаттары және басқалар кіреді. Барлық мәтіндер Гонконг қор биржасынан, Гонконгтың бағалы қағаздар және фьючерстер комиссиясынан және Гонконг үкіметінің веб-сайттарынан алынған. Жинақ дереккөз түріне сәйкес ұйымдастырылған және .tsv. файл форматына сәйкес келеді.

COCA – Corpus of Contemporary American English. 1990-2015 жылдар аралығында 520 миллионнан астам сөз жиналған, корпустар ағылшын әдет-ғұрыптарын ауызша сөйлеу, көркем әдебиет, жаңалықтар көздері мен ғылыми көздерден алынған [76]. Корпус 485 202 мәтінде қолданылған 1 миллиардтан астам сөзден тұрады, оған 1990-2019 жж. аралығында жыл сайын 24-25 миллион сөз қосылып отырған. Корпус жыл сайын теледидар бағдарламалары мен фильмдер субтитрлері, ауызша тіл, көркем әдебиет, журналдар, газеттер және академиялық журналдар жанрлары арасында бірдей пропорцияны сақтап отырады [77].

1810 жылдан 2009 жылға дейінгі кезеңде жиналған 400 миллионнан астам сөзі бар корпус (COHA – Corpus of Historical American English) тарихи ағылшын қолданыстары мен грамматикалық құрылымдарды іздеуге мүмкіндік береді.

AnCora Catalan and Spanish Corpus. AnCora деген Catalan (AnCora-CA) және Spanish (испан) корпусынан (AnCora-ES) тұрады, олардың әрқайсысы 500 000 сөзді құрайды. Корпустарға әртүрлі деңгейде аннотациялар берілген.

1200-1900 жылдар аралығында испан тіліндегі мәтіндерді Corpus del Español-дан іздеуге болады. Жақында корпус кеңейіп, 1900-2000 жылдардағы мәтіндерді қамтитын болды.

CREA – Corpus de Referencia del Espanol Actual – Корольдік испан академиясының бұл мәтіндік корпусында бүкіл испан тілді әлемнің кітаптары мен мақалалары бар.

Ал CU Linguistics' Verb Corpora Server CU тіл білімі бөлімінде өз корпус-тары үшін интернетте «тілі бойынша», «аты бойынша» іздеу тетігі бар сервер. Кейбір деректер дискілерде немесе басқа форматтарда сақталған.

Соңғы екі онжылдықтағы Париж тұрғындарынан алынған сұхбаттар жиынтығы. Аудиофайлдар мен стенограммалар жүктеп алушылар үшін қолжетімді корпус Discours sur la ville Corpus de Français Parlé Parisien des années 2000 (Corpus of Parisian French from the 2000's) деп аталады.

Корпустарды іздестіру қызметін көздейтін орталықтандырылған сервистік корпус түрі ELRA Catalogue of Language Resources деп аталады. ELRA каталогы тілдік корпус-тың бар-жоғын анықтауға және оларға қол жеткізуге көмектесетін тілдік ресурстар қорын ұсынады. Бұл корпустардың кейбіреулері қосымша тіркелуді қажет етеді.

Француз тіліндегі корпустар (French Language Corpus Collection), мәтіндік және аудио файлдар, ересектер мен балалардың өзара сөйлесім қарым-қатынасына, сондай-ақ ересектердің өзара коммуникациясына ерекше назар аударады.

Ағылшынтілді ортақ халықаралық интернет-ресурс International Corpus of English (ICE) деп аталады және бұл жерде орналасқан әрбір ICE корпусы 1989 жылдан кейін Канада, Шығыс Африка, Ұлыбритания, Гонконг, Үндістан, Ирландия және Австралия, Ямайка, Жаңа Зеландия, Нигерия, Филиппин, Сингапур, Шри-Ланка және АҚШ корпустарымен жиналған 1 миллион жазбаша және ауызша ағылшын тілінің қорынан тұрады.

Мичиган университетінде академиялық спектакльдердің транскриптітерінің онлайн жиынтығы (сонымен қатар кейбір сөйлеу аудиожазбаларын қамтуы мүмкін) ғана арналған мынадай MICASE - Michigan Corpus of Academic Spoken English корпус түрі де бар екен. Бұл корпус-ты ғылым тілінде сөйлеу сипаттамаларын зерттеу үшін ғылыми жобаларда пайдалануға арналған.

Тарихи ағылшын тілін зерттеу корпусын Penn Parsed Corpora of Historical English ұлттық корпус-тың тарихи ішкорпусы деп те бағалауға болады, оның ішінде орта ғасырдағы ағылшын тілін талдау бойынша Пенн-Хельсинки корпусы, екінші басылым (PPCME2), ерте немесе заманауи ағылшын тілін талдау бойынша Пенн-Хельсинки корпусы (PPCEME) және қазіргі Британдық ағылшын талдау бойынша Пенн корпусы, екінші басылым (PPCMBE2),

Британдық ағылшын прозасының мәтіндері мен мәтіндерінің үлгілері оның бүкіл тарихында – ең ерте орта ағылшын тілінен Бірінші дүниежүзілік соғысқа дейінгі құжаттар қамтылған. Мәтіндер үш формада ұсынылған: қарапайым мәтін, сөз таптары мен синтаксиске арналған мәтіндер.

Лос-Анджелес қаласындағы Калифорния университетінің фонетика зертханасы ғалымдары ана тілінде сөйлеушілер мен тіл үйренушілер үшін құнды болып табылатын фонетикалық және фонологиялық зерттеулер үшін әлемнің жүздеген тілдерінің жазбаларын жинаған. Осы сайттағы материалдарда 200-ден астам тілде фонетикалық құрылымдарды суреттейтін аудио жазбалар UCLA Phonetics Lab Archive корпусына жинақталған [76].

TS Corpus: online corpus of Turkish – TS Corpus: түрік тілінің онлайн корпусы. Түрік корпусын құру жобасы. Құрамында миллиардтан астам POST тарға негізделген жинақ. TS Corpus құрылысын (Taner мен Turker Sezer) Танер мен Түркер Сезер жүргізеді [78].

Ағылшын тілінің корпусына дүние жүзіндегі жүздеген университеттер мен мыңдаған жеке тұлғалар, сондай-ақ Германияның зерттеу қоры сияқты ұйымдар қолдау көрсетеді. Корпус деректерін (мысалы, толық мәтінді, сөз жиілігі жөніндегі ақпаратты) Amazon, Google, Microsoft, IBM, Sony, Disney, Intel, Adobe, Samsung деген сияқты көптеген компаниялар, бірнеше (AI) жасанды интеллект компаниялары, Merriam-Webster, Dictionary.com, Grammarly, Duolingo, TurnItIn, Oxford University Press, Sketch Engine сияқты тікелей тілге қатысты компаниялар, т.б. пайдаланып келеді.

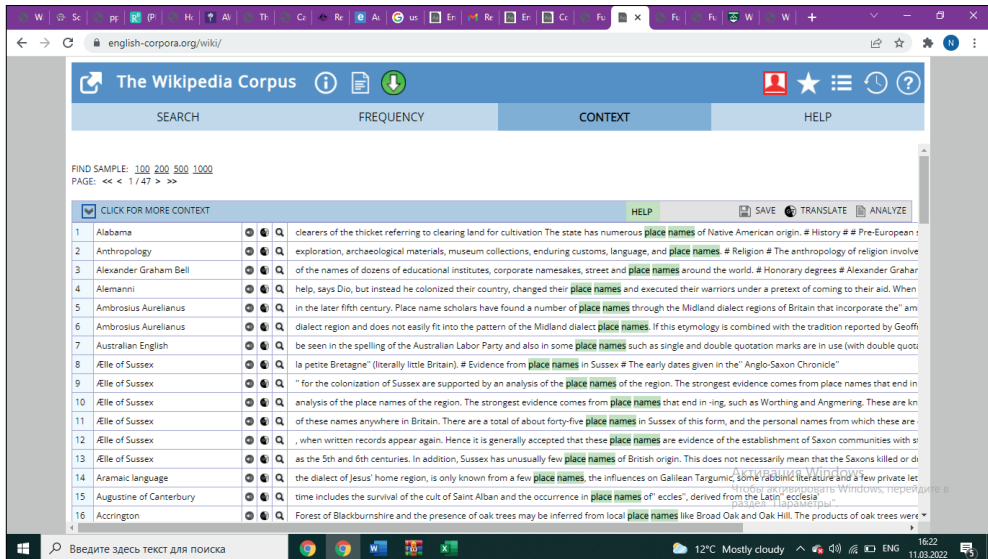
Бұл корпус төмендегідей бірнеше субкорпустардан тұрады:

- News on the Web (NOW) Ғаламтордағы жаңалықтар
- iWeb: The Intelligent Web-based Corpus – Зияткерлік вебке негізделген корпус
- Global Web-Based English (GloWbE) – Ағылшын тіліндегі ғаламдық веб-сайт
- Wikipedia Corpus – Уикипедия корпусы
- Coronavirus Corpus - Коронавирустық корпус
- Corpus of Contemporary American English (COCA) – Қазіргі америкалық ағылшын тілі корпусы (COCA)
- Corpus of Historical American English (COHA) – Тарихи америкалық ағылшын тілі корпусы (COHA)
- The TV Corpus – Теледидар корпусы
- The Movie Corpus – Көркем фильмдер корпусы, диалектілер: 6 ел
- Corpus of American Soap Operas – Америкалық сериалдар корпусы
- Hansard Corpus – Хансард корпусы
- Early English Books Online – Ерте ағылшын тілінің онлайн кітаптары
- Corpus of US Supreme Court Opinions – АҚШ Жоғарғы соты пікірлерінің корпусы
- TIME Magazine Corpus – TIME журналының корпусы
- British National Corpus (BNC) – Британ ұлттық корпусы (BNC)
- Strathy Corpus (Canada) – Страси корпусы (Канада)
- CORE Corpus – CORE корпусы

- American English – Америка ағылшыны
- British English – Британ ағылшыны [79].

Осы корпустардың бірнешеуіне ономастика саласына қатысты 'onomastics', 'geographical names', 'place names', 'exonym', 'endonym' терминдерін енгізгенде, төмендегідей нәтижелер шықты:

- 1) WIKIPEDIA CORPUS-нда (6-сурет):
 - 'onomastics' – 198
 - 'geographical names' – 645
 - 'place names' – 4667
 - 'exonym' – 491
 - 'endonym' – 383



Сурет 6. WIKIPEDIA CORPUS

- 2) COCA Қазіргі америкалық ағылшын тілі корпусында:

- 'onomastics' – 19
- 'geographical names' – 3
- 'place names' – 317
- 'exonym' – 0
- 'endonym' – 0

- 3) BRITISH NATIONAL CORPUS Британ ұлттық корпусында (BNC):

- 'onomastics' – 2
- 'geographical names' – 1
- 'place names' – 63
- 'exonym' – 2
- 'endonym' – 5

- 4) NOW CORPUS Ғаламтордағы жаңалықтар корпусында:

- 'onomastics' – 48

‘geographical names’ – 451

‘place names’ – 4353

‘exonym’ – 22

‘endonym’ – 14

5) THE TV CORPUS Теледидар корпусында:

‘onomastics’ – 0

‘geographical names’ – 0

‘place names’ – 36

‘exonym’ – 0

‘endonym’ – 0

Бұдан көретініміз, ағылшын тілінің ішкорпустарында топонимикалық мәліметтердің қамтылу аясы әркелкі. Демек, ономастика саласы терминдерінің мұндай әр корпуста түрлі жиілікте кездесу нәтижесі, тиісінше, әр корпустың мүмкіндігіне қарай өзінің тақырыптық аясына байланысты жұмыс істейтінін көрсетеді.

Келесі көлемді корпустың бірі Орыс тілінің ұлттық корпусы. Орыс тілінің Ұлттық корпусы – 500 миллионнан астам сөздерді қамтитын мәтіндердің ең үлкен электронды жинағы. Бұл орыс тілін зерттейтін лингвистердің негізгі көздерінің бірі. Жоба 9 жыл бұрын, 2004 жылдың сәуірінде іске қосылды. Корпус көптеген ұйымдардың, соның ішінде РҒА В.В.Виноградов атындағы Орыс тілі институты, РҒА ақпарат беру проблемалары институты, ММУ және басқалардың күшімен толықтырылып, қолдау көрсетіледі.

Орыс тілінің Ұлттық корпусы қазіргі кезде келесі шағын корпустардан (субкорпустардан) тұрады:

- әрбір сөйлемге толық сипаттағы морфологиялық және синтаксистік құрылымдары (тәуелділік сызбасы) (деревозависимость) салынған терең аннотацияланған корпус;

- орыс немесе ағылшын тіліндегі белгілі бір сөздің не сөз тіркесінің барлық аудармаларын іздеп табуға мүмкіндік беретін мәтіндердің орысша-ағылшынша параллель корпусы;

- Ресейдің әртүрлі аймақтарының диалектілік сөйлеу тілінің жазбасын жадында сақтайтын және олардың грамматикалық ерекшеліктерін ескере отырып, диалектілік морфология бойынша арнайы іздеу үдерісі қарастырылған диалектілік мәтіндер корпусы;

- тек лексикалық және грамматикалық белгілер ғана емес, сонымен бірге өлеңдерге тән белгілер бойынша іздеу (мысалы, сонета, эпиграмма, өлеңдер және т.б. белгілі тіркестерді іздеу) мүмкіндіктері қарастырылған поэтикалық мәтіндер корпусы;

- орыс тілінің мектеп бағдарламасына бағытталған белгіленімі бар, омонимдер қатары ажыратылған корпус – орыс тілінің оқыту корпусы;

- көпшілік және жекелік сөйлеу тілінің, сонымен бірге 1930-2000 жылдардағы кинофильмдер транскриптерінің магнитофондық жазбаларының мағынасы ашылған орыс тілінің сөйлеу тілі корпусы [12, 46].

Бұл корпуста ономастикалық атаулар лингвистикалық ақпаратпен қамтамасыз етілген деуге болады. Біздің корпусымыз бастапқыда топонимдер бо-

йынша құрастырылып жатқандықтан, орыс корпусындағы топонимдердің аннотациялануына назар аудардык. Бірінші, табиғи нысан жайлы геоақпаратты, атаудың мәдени семантикасы мен этимологиясын, орфографиясын көрсетілген сілтеме арқылы өтіп, academic.ru сөздігінен көруге болады. Мысалы, Волга жалқы есімі жайлы сөздіктен 200 түсініктеме табылды. Ол дереккөздерден атаудың семантикасына кіретін ақпараттың кең көлемін табуға болады. Екінші, жалқы есімнің қай түрі болуына байланысты, жер-су атауы болса *t:topon*, *r:propn* деп, ал кісі аты болса *t:persn*, *r:propn*, *t:hnm* деген шартты белгілермен таңбаланған. Орыс корпусында топонимдерінің басым бөлігі аннотацияланғанына көз жеткіздік.

Біз, ең алдымен, ономастикалық/топонимдік ішкорпусты құрастыруда топонимдерді сипаттау критерийлерін әзірледік. Мәселен, жер-су атауларының корпусын құру әр түрлі жанрдағы көптеген мәтіндермен, көркем әдебиеттермен, публицистикамен, бұқаралық ақпарат құралдарындағы мәтінмен жұмыс жасауды қамтитыны белгілі. Мұндай жұмыс атаулардың контекстік ортасын бекіту және оның семантикалық көлемін сипаттайтын барлық мүмкін жағдайларды анықтау үшін маңызды саналады. Мұндай мәтіндерді жүйелеудің міндетті шарты – олардың төлқұжатын құру, яғни басылымның толық деректерін, автордың жасын, жынысын көрсету және т.б.

Сондай-ақ, топонимдердің төлқұжатына төмендегідей белгіленімдер енгізіледі (7-сурет):

- 1) Географиялық нысанның типі;
- 2) Топонимнің орфографиялық нормасы мен бейнормасы. Мәселен, Қарағаш/Қарағаш, Сарыарқа/Сарарқа және т.б.;
- 3) Топонимнің сөзжасамдық ерекшеліктері;
- 4) Топонимнің хронологиялық мәліметі;
- 5) Ареалы мен географиялық орналасуы;
- 6) Этимологиялық және лексика-семантикалық анықтамасы

	A	B	C	D	E	F	G	H	I
1	Онім түрі Белкайкаң	Шартты белгісі t:topon	Топоним түрі ойконим	Шартты белгісі t:oiKon	Орфографиялық норма Белкайкаң	Орфограмма (Бейнорма) жоқ	Сөзжасамы Бел+кайкаң (зат есім + зат есім) (*Жер-су аттарының сөзжасамдық үлгілері* 261 б.	Геоақпарат (кітап) БЕЛҚАЙҚАҢ - қоныс Итжон үстіртінің оңт. бөлігінде, Талғар ауданда (*Жерінің аты, елінің хаты* 160 б.)	Геоақпарат (инте) https://kk.wikipedia.org/wiki/%D0%98%D1%8B%D0%BE%D1%8F
2	Белкайың	t:topon	Ороним	t:opon	Белкайың	Белкайың, белкайың	Бел+кайың (зат есім+зат есім) (*Жер-су аттарының сөзжасамдық үлгілері* 261 б.	БЕЛҚАЙЫҢ - жота. Тастау өз нін оң жағалауында, Алақон ауданда. (*Жерінің аты, елінің хаты* 160 б.)	https://kk.wikipedia.org/wiki/%D0%99%D0%9B%D2%9B%9D%D0%B9%D1%8F A3 https://stom.tlimes.ta-shi-oleshop-ami.januarlardan-elik-araci-a.html?pa
3	Белқұдық	t:topon	гидроним	t:gidron	Белқұдық	елқұдық, белқұдық, белқұдық	Бел+құдық (зат есім+зат есім) (*Жер-су аттарының	Белқұдық - Бақалас ауылының солтүстік-шығысқа, Бірік ауылының	деректер жоқ

Сурет 7. Ономастикалық корпустағы топонимдер төлқұжаты

ҚТҰК-ның ономастикалық ішкорпусын жасауда осында (7-суретте) көрсетілген белгіленімдер топонимдер төңірегіндегі ең маңызды ақпарат ретінде танылып, біздің ішкорпус жасаудағы басты ұстанымдарымызды құрап отыр.

Көпшілік ғаламтор желісін ақпарат алу үшін қолданатыны белгілі. Әдетте адамдар бір немесе бірнеше тірек сөзді тереді, компьютер осы сөздерге байланысты бар ақпарат көздерін немесе арнайы сайттарды тауып береді. Бұл ақпарат алудың бір тәсілі. Берілген сөзге қатысты ақпаратты табу үшін лингвистикалық сараптама қолданылады. Онда тірексөздердің қолданылу жиілігіне, морфологиялық формаларына, мағыналас сөздеріне, семантикалық жақтан сәйкес ұғымдарға байланысты ғаламтор сайттар жауап береді. Сол арқылы керекті мәліметті ұсынады» [80, 519-520]. Алайда ғаламтордағы мәтіндер мен тілдік корпустағы мәтіндердің айырмашылығы бар. Бұл туралы Э.Д.Сүлейменова «Цифрлендірілген мәтіндердің базасы корпус белгісінің көмегімен арнайы өңделеді, ал электронды кітапханалар мұндай өңдеуден өтпейді. Дәл осы белгі корпусты түбегейлі тілдің қолданылуы мен өзгеруі жайлы ақпараттың жаңа және объективті дереккөзіне айналдырады. Белгі (аннотация, код) – мәтіндердің және олардың бірліктеріне тән қасиеттер жайлы арнайы лингвистикалық ақпарат беретін корпустың негізгі сипаттамасы. Корпустың белгісі дәлірек және әрқилырақ болған сайын оның әлеуметтік, ғылыми және оқыту құндылығы жоғарырақ болады. Кейбір корпустар аз ғана белгімен қанағаттанса, кейбіреулерінде белгі ұдайы жетілдіріліп отырады» [81, 77]. Мәселен, Google-ге Жетісу деп жазсақ, ол туралы қыруар материалдар шығады. Алайда ол жерде Жетісу атауы туралы лингвистикалық (ономастикалық) ақпарат берілмейді. Ономастикалық ішкорпустағы белгіленімдердің түрлеріне қарай төмендегіше түсіндірме беруге болады:

а) топонимнің түрі: ороним, гидроним, ойконим, некроним т.б. байланысты шартты белгілер енгізіледі (3-кесте).

Кесте 3

Топонимдер түрлері мен анықтамалары

Топонимнің түрлері	Түсініктемесі	Шартты белгілері
ойконим	Елдімекен атаулары: қала, ауыл, аудан, қоныс, қыстау, кент, мекен,	t.oikon
ороним	Жер бетінен жоғары және төмен жатқан жер бедері атаулары: Тау, төбе, кезең, жыра, сай, мүйіс, шоқы,	t.oron
гидроним	Су нысандарының атаулары: өзен, көл, бұлақ, тоған, канал, құдық т.б.	t.gidron
хороним	Аймақ, өлке атауы	t.horon
некроним	Жерлеу орындары: мола, бейіт, мазар, қорым	t.nekron
дримоним	Орман, алқап, тоғай атаулары.	t.drimon
дромоним	Кез келген байланыс, қатынас жолының, жол тораптарының атауы	t.dromon

ә) атаудың құрылымына байланысты сөзжасамдық белгіленімдер енгізіледі. Топонимдердің сөзжасамдық ерекшеліктері қарастырылады. Әрбір атаудың ең алдымен қандай сөзден немесе сөздерден тұратыны, оның құрамы мен қай сөз табынан жасалғаны көрсетіледі. Мысалы, Сарыарқа – Сары + арқа, күрделі жалқы есім: біріккен сөз – сын есім + зат есім. Зат есімдердің белгіленімдерін орнатудың үлгісін А.Жұбанов пен А.Жаңабекова төмендегідей көрсетеді. «1. Түбір морфема зат есім; 2. Туынды зат есім (түбір + жұрнақ); 3. Күрделі зат есім: а) біріккен сөз; ә) қосарланған сөз; б) құрама сөз; қысқарған сөз» [12, 112]. Б.Бияров жер-су атауларының сөзжасамдық үлгілерін зерттеген еңбегінде топонимдердің жасалу модельдері бойынша топтайды. 1. Түр-түс мәнді сын есім + зат есім үлгісімен жасалған біріккен топонимдер (Ақайдар, Алакөл, Бозарал, Жиренкөл, Көкбел, Қарақұдық, Қоңыртөбе, Қызылағаш); 2. Сапа белгісін білдіретін сын есім + зат есім үлгісімен жасалған біріккен топонимдер (Айырбидайық, Арықбалық, Байсала, Баянауыл, Беріктас, Бособа, Ескікеңсе, Жаксықұм, Жалтыржон, Жаңадала); 3. Көлемін және бағытын білдіретін сын есімдер + зат есім (Бертоған, Биіккөл, Дәутөбе, Дөңгелеккөл, Жалпақбас, Жіңішкебұлақ, Кішкенесай, Келтетал, Кішікемер, Қысқабұлақ); 4. Зат есім + Зат есім үлгісімен жасалған біріккен топонимдер (Кемербастау, Кепесай, Қабанасу, Науатоған); 5) Сан есім + зат есім үлгісімен жасалған біріккен топонимдер (Алпыссала, Бесарық, Алтыағаш, Егізтөбе, Жартыкөл, Жүзқарағай, Тоғызак) [82].

б) орфографиялық нормасы мен бейнормасы беріледі. Әлеуметтік желіде, түрлі баспасөзде, тағы да басқа дереккөздерде жер-су атауларын жазуда орфографиялық ала-құлалық туындап жатады. Бұл белгіленім түрі топонимдерді біріздендіруде үлкен рөл атқарады.

Ономастикалық ішкорпусты құрастырудың маңызы жоғары және оны құрастыру өте ауқымды жұмыс екеніне көз жеткіздік. Бұл тақырыпта әлемдік корпустардағы ономастикалық атаулардың белгілену ерекшеліктерін анықтап, өз тәжірибемізде ұтымды пайдалану жолдарын қарастырдық. Орыс тілінің ұлттық корпусындағы мәліметтерді талдай отырып, олардағы электронды дереккөздердің мол екенін байқадық. Бұл өз кезегінде корпустың көлемді, толық болуындағы негізгі шешуші фактор болып табылады. Сондықтан қазақ тілінде жарық көрген барлық энциклопедия, сөздік, анықтамалықтардың электронды түрде қолжетімді болуы Қазақ тілінің ұлттық корпусын құрастыруға үлкен септігін тигізетін болады.

Сонымен, қазақтың ежелгі тарихынан бүгінгі күнге дейін келе жатқан жер-су атауларымыздың корпуста көрініс табуы, олардың метабелгіленімдерінің айқындалуы ғылыми-көпшілік қауымға пайдасы үлкен болады. Ономастикалық ішкорпусты географ, тілші, тарихшы, этнограф мамандар, саяхатшылар, университет және мектеп оқытушылары мен оқушылары, сондай-ақ көпшілік қауым пайдалана алады. Өйткені кез келген ізденуші ономастикалық атаулардың үлкен қоймасына тап болады. Онда ақпараттар топтастырылып, жүйелі түрде берілуі ізденушіге керек мәліметті жылдам табуға көмектеседі әрі жан-жақты білім беретін болады. Алдағы уақытта ҚР барлық облыстарының топонимдері ономастикалық ішкорпусқа енгізіледі. Жай ғана енгізіліп қоймай, әр жер-

су атауының мәдени семантикасы, мүмкіндігінше этимологиясы да беріліп отырады. Яғни бұл ұлт игілігіне қызмет ететін ауқымды заманауи дерекқор көзі болмақ. Сонымен бірге ономастикалық белгіленімдер де кеңейіп, жетілдіріліп отыратын болады және келешекте әр атауды интернет желісіне нақты географиялық нысанның 2GIS іздеу-орналасуын көрсететін сервисі арқылы геоақпараттық сипаты да берілмек.

3.3 Қазақ тілінің ономастикалық ішкорпусын әзірлеу тәжірибесі

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

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 корпусстың 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).

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 тандалған 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 [83, 17] ізімен ономастикада қалыптасқан төмендегідей негізгі топтарда қарастырамыз:

ороним – жер қыртысының бетіндегі, жер бедері (ландшафт, рельеф) шығыңқы (положительный рельеф) бедері → тау, қыр, төбе, шоқы, дөң, бел, 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, қаяқ, бұлақ, жылға, қайнар, бастау, бөген, саз, батпақ.

ойконим – *астионим* қала атауы, *комоним* – кез келген ауыл типтес елді мекен, *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 ішкорпустан 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 кеңейіп түседі. Мысалы, топонимдер тіл білімінің ғана емес, география ғылымының да зерттеу нысаны болып саналады. Сондықтан ОІК-та географ мамандар да 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-да 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.

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, диалектілік, мәдени, мақал-мәтел, тарихи, параллель, жарнама және ономастикалық ішкорпус* іске қосылды. Әр ішкорпус өзіне тән мазмұнымен, соған сәйкес белгіленім түрлерімен ерекшеленеді. Ономастикалық ішкорпуста басқа корпустардан даралайтын бірқатар өзгешеліктері бар. ОІК осы күнге дейін қалыптасқан академиялық қазақ ономастикасына табан тірейді. Ішкорпуста пайда болуы, қазақ ономастика ғылымының қалыптасып, дамуымен тығыз байланысты. Ономастикалық зерттеу нәтижелері ономастикалық корпуста 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.

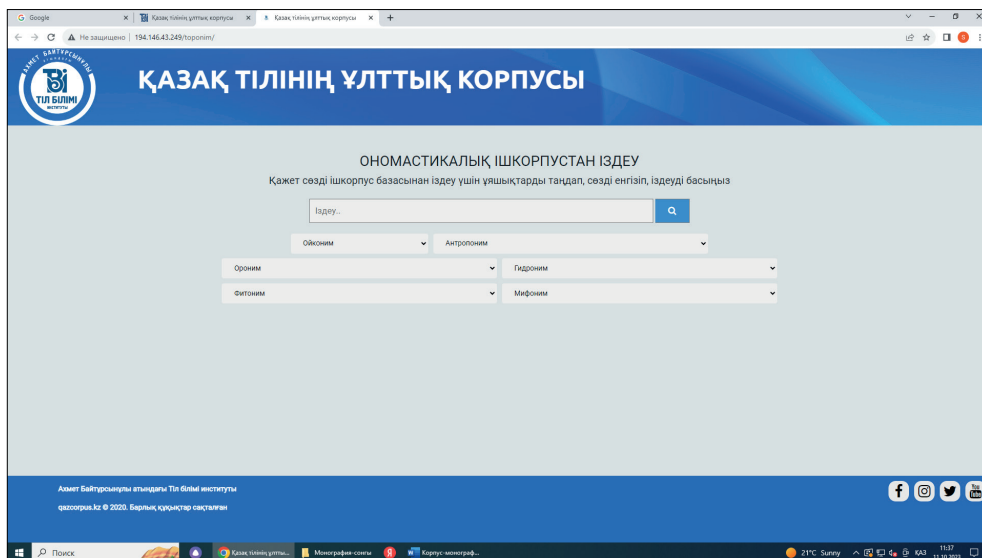
2) 1970-1995 жж. арасы. Бұл кезеңде (ономастикалық) аймақтық топонимиялық зерттеулер дами түсті. Ономастикалық теорияларға назар аударылды. Топонимдер мен антропонимдер этнолингвистикалық бағытта зерттеле бастады. Ономастикалық зерттеулердің ауқымы кеңейді. Космонимдер, зоонимдер, Қазақстан территориясындағы 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 антропоэзеттік 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.Сұлтанъев, Е.Ә.Керімбаев, В.У.Махпиров, Б.Бектасова, М.Қожанов, Г.Б.Мәдиева, Қ.Рысберген, А.Жартыбаев, Ғ.Сағидолда, Б.М.Тілеубердиев, К.М.Головина, М.А.Диарова, С.Иманбердиева, С.Е.Керімбаева М.Е.Какимова т.б. [84, 8].

ҚТҰК-да жұмыс істеу барысын көрсетсек (8-сурет). Алдымен <https://qazcorpus.kz/news/> [40] сайтына кіресіз. Корпус ұлттық корпус болғандықтан ақпараттар қазақ тілінде берілген. Басты бетте «Іздеу» тармағы тұрады, оны бассаңыз ұлттық корпусың ішіндегі 10 субкорпусты көрсетеді. Солардың қатарында «Ономастикалық ішкорпус» тұрады. Сол парақшаға өтсеңіз, төмендегі бетті көресіз.

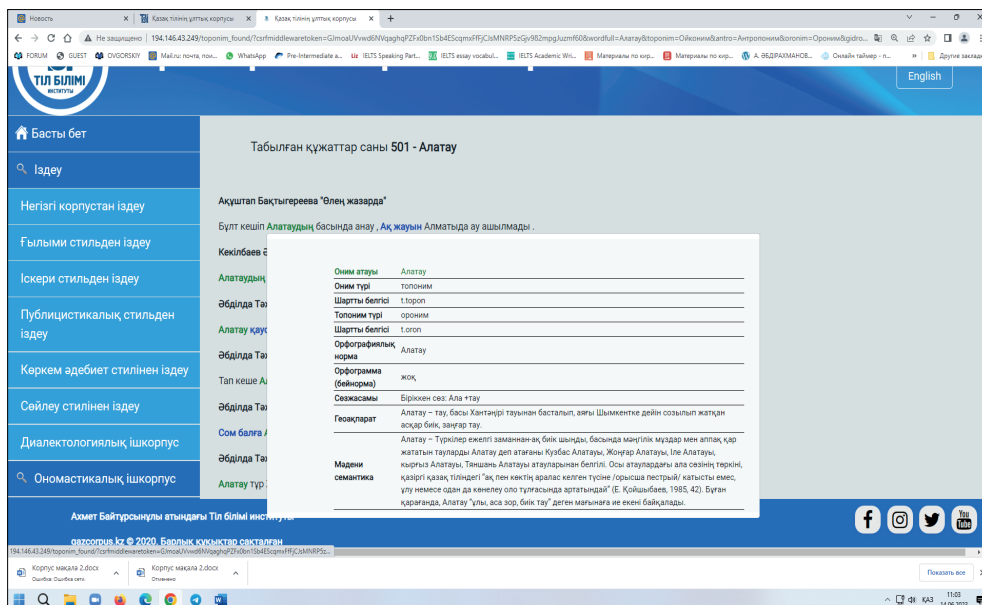


Сурет 8. Ономастикалық ішкорпустың беткі беті

Ұлттық корпусың бір бөлігі саналатын ОІК-да 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 көрсетілетін болса (9-сурет), шығарма авторының жаңына апарсақ, автор мен шығарма туралы толық ақпараты беріледі (3-сурет).



Сурет 9. Онимнің лингвистикалық белгіленім терезесі

а) *Шартты белгілер*. Жалқы есімнің түрлеріне сәйкес шартты белгілер қойылды. Мысалы, топоним – t.toron, антропоним – a.antoron, 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*. ОҚ-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://qazcorp.kz/>) [40].

Проф. Г.Б.Мәдиева бұл ақпараттар төмендегідей мағлұматтармен толықтай түсуді қажет деп табады:

– Оронимдер мен гидронимдер үшін маңызды саналатын географиялық сипаттамасының берілуі (биіктігі, ұзындығы, көлемі және т.б.)

– Ойконимдер мен урбанонимдер үшін маңызды саналатын хронологиялық ақпараттар [85, 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. Севортьян, Г. Дерфер, Г. Рамстедт, Г. Вамбери, А. Щербак, А. Әбдірахманов, Ғ. Қоңқашпаев, Т. Жанұзақов, т.б.). Оның ішінде Ғ. Қоңқашпаевтың келтірген дерегі дұрыс деп есептейміз. Моңғ. алт (алтын)+тай (туынды сын есім жұрнағы), сонда Алтай>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 [86, 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 үйренегіндерге, сондай-ақ дәстүрлі ұлттық білімдер жүйесіне қызығушылық танытқандарға, жастарға қазақ онимдерінің мәдени семантикасын, ұлттық кодты меңгеруге көмектеседі.

д) *Метабелгіленім*. ҚТҰК 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 жиған, әрі өзгеге жеткізе отырып, белгілі бір уақыт пен кеңістіктен мәлімет (дерек) беретін, жеке тұлғаның рухани әлемін ашуға ықпалы зор күрделі тілдік таңба» [87, 8] деген болатын.

Корпустың тірегі – мәтін. Мәтін сапалы, көп болған сайын корпус та жақсы болады. Корпусқа кез келген мәтінді ала алмаймыз. ҚТҰК-ның негізін қалаған, 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 корпустың мақсатына сай белгілі бір стильдер мен жанрлар немесе кезеңдер бойынша теңгерімді мөлшерге сай болуы қажет. Яғни мәтін де жасалып отырған корпус түріне қарай сәйкестендіріледі» [12, 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» [88, 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 корпустың зерттеу мақсатына қарай көлемді мәтіндер іріктеліп алынатындағын айтады. Корпус дегеніміз компьютер жадына сақталған көлемді мәтіндер жинағы, ал корпустық лингвистикада ауқымды электронды мәтіндерді компьютердің көмегіне сүйене отырып зерттейді. Сол себептен жас ғалым корпустық лингвистика мен мәтін лингвистикасының өзара тығыз байланысты екеніне назар аударады [89, 303].

Енді ономастикалық ішкорпустың 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. 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. Әртүрлі 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» [12, 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» [90] деп, орыс ғалымы да айтқан болатын.

Корпуста «метабелгіленім» термині қолданылады. Белгіленім корпустаң 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 ішкорпустың метабелгіленімі төмендегі кестеге сәйкес қойылды (4-кесте).

Кесте 4 – Метабелгіленімі қойылған мәтін

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 корпуста 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 ішкорпуста 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 ішкорпустаң 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.

Тақырыпта 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 ОИК-қа антропонимдер мен топонимдер енгізіліп, оларға тілдік белгіленім жасалды. Ономастикалық ішкорпусқа базасын қазақ ономастикасы бойынша қорғалған диссертациялар мен жарық көрген монографиялар құрайды. Мақалада осы дереккөздерге жасалған метабелгіленімнің үлгісі берілді. Бұл белгіленім түрі негізгі корпус пен барлық ішкорпустарға ортақ тақырыпшалардан тұрады.

ОИК-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, толығырақ 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 қалыптануы 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.

3.4 Мифонимдер мен теонимдерді Қазақ тілінің ұлттық корпусына енгізу технологиясы

Инновациялық компьютерлік технологиялар бізге Интернет жүйесі арқылы деректердің айтарлықтай көлемін, соның ішінде қарапайым адамдар жазған мәтіндерді жинақтауға және оларды талдауға, белгілі бір сөздің ресми құжаттардан бейресми мәтіндерге дейінгі кең ауқымдағы мәтіндерде қалай қолданылатынын анықтауға мүмкіндік береді. Осы орайда біз корпусық лингвистикаға сүйенеміз. Корпусық лингвистика – тілді ақпараттандыру, техника мен технологияны пайдалана отырып зерттеуді қажет ететін маңызды ғылым саласы. Корпусық лингвистика лингвистикалық корпусарды құрастырып, оны пайдаланудың ұстанымдарын зерттейді [12].

Әр тілдің ұлттық корпусы белгілі дәрежеде Ұлттық тілді толық түрде бейнелейді. Ұлттық корпусар әр тілдің тілдік заңдылықтарын, оның ішінде тұрақты лексика-грамматикалық жүйесін толық қамтыған миллиондаған сөз-қолданыстан тұратын электронды пішіндегі көлемді мәтіндер жинағынан тұрады. Ол тілші-ғалымдарға, мектеп оқушыларына, студенттерге, мектеп және жоғары оқу орындарының оқытушыларына, басқа да сала мамандарына түрлі зерттеулер жүргізуге, танымы мол ақпарат алуға қажетті мүмкіндік береді.

А.Байтұрсынұлы атындағы Тіл білімі институты Қазақ тілінің ұлттық корпусын құрастыру мәселесі бойынша 2009 жылдан бастап зерттеу жүргізіп келеді. Қазақ тілінің 5 стилі бойынша (көркем әдебиет, публицистикалық, іскери, ғылыми) интернет желілерінен, электронды кітапханалардан, баспалардан т.б. дереккөздерден мәтіндер жинақталып, оларға 23 параметрден тұратын мета-белгіленім (мәтіннің авторы, автордың жасы, мәтін тақырыбы, жазылған стилі, жанры, мәтін типі, дереккөзі т.б.) енгізіліп, түрлі белгіленім (морфологиялық, лексикалық, мәдени-семантикалық) қойылды. Қазіргі уақытта ҚТҰК мәтіндік базасының көлемі – 30 миллион сөзқолданыстан тұрады [40].

Лингвистикалық корпус, яғни мәтіндер корпусы – бірізденген, бірыңғайланған, шартты белгілері қойылған тілдік деректер болып саналады. Корпусқа мәтіндер арнайы бағдарламалар арқылы стилі анықталып, тиісті мәліметтер енгізіледі. Корпусқа енгізілген мәтіндердегі әрбір тілдік бірлікке, ол не жеке сөз не қосымша болсын, лингвистикалық ақпарат беріледі. Тілде мұндай лингвистикалық ақпарат беруді аннотациялау деп атайды. Мәтіндер бірліктеріне берілген лингвистикалық ақпараттардың мол болуы осы белгіленімдердің әртүрлілігіне байланысты. Белгіленімдер тілдегі жекелеген деңгейлерге тән тілдік мәліметтерді қамтиды. Қазіргі Қазақ тілінің ұлттық корпусында морфологиялық, синтаксистік, мәдени-семантикалық белгіленімдер бар. Бұл белгіленімдер сөзді түбір мен қосымшаға ажыратуда, сөйлем талдауда аса қажет. Ал мәдени-семантикалық, лексика-семантикалық белгіленімдер қою, жетілдіру жұмысы – негізгі өзекті мәселе болып отыр.

Компьютерлік технологияның дамуы әртүрлі мақсаттар үшін зерттеушілердің сұраныстарын қанағаттандыра алатын мәтіндердің әртүрлі типтерін құруға ықпал етті. Осылайша, қазіргі уақытта көп мақсатты, мамандандырылған, жазбаша, ауызша, аралас, бір тілді, көп тілді, параллель, әдеби, диа-

лектілік, терминологиялық, ауызекі, фольклорлық, драмалық корпустар және т.б. бар. Жоғарыда аталған корпустар лингвистер әзірлеген, содан кейін бағдарламаланатын ережелер негізінде жүргізілді. Мәтіндерді өңдеу технологиялары мен жүйелерін дамытудың қазіргі кезеңі статистикалық қағидат негізінде корпустар үшін деректерді жинауға және өңдеуге мүмкіндік береді: Интернет желісіндегі мәтіндік фрагменттерді арнайы іздеу бағдарламаларымен салыстыру, талдау және ең жақсы сәйкестіктерді таңдау арқылы.

Соның ішінде мифонимдер мен теонимдер белгіленімдерін жасау және енгізудің негізгі мақсаты – қазақ тілінің ұлттық корпустық базасын жасау; базаға мифтік және діни тілдік таңбаларды жинақтау; мағыналарын ашу, белгіленім қою, түсіндіру, мысал жинау. Зерттеуіміздің негізгі көзделіп отырған мақсаты да осы ономастикалық ішкорпустағы мифонимдер мен теонимдердің мәдени кодын анықтап белгіленім қою.

Қазақтың ұлттық таным көкжиегінде мифология мен теологияның алар орны зор дейтін болсақ, ол ғылым саласы ретінде бүгінгі таңда енді-енді бойын тіктеп келе жатыр. Біз бүгінгі ғылым арнасында мифология мен теологияны соның ішінде дербес ғылым саласы ретінде танып, оған жаңаша көзқарас танытып отырғандықтан, түп тамырында үш бірліктің жатқанын айта кетуіміз жөн. Олар: наным, дін, философия.

Адамзаттың көне замандардағы шынайы қоршаған ортаны бейнелеудегі мифологиялық, теологиялық ойлау қызметінен әлемнің мифологиялық және теологиялық бейнесі қалыптасып, адам табиғатты, әлемді мифологиялық және діни тұрғыдан тану деңгейіне көтерілді.

Адамдардың мұндай тылсым табиғат сырларына байланысты дүниетанымын профессор Қ. Жұбанов былайша түсіндіреді: «Адамдар бұрын табиғат сырының күштерін білмеген. Сондықтан олар табиғатқа адамша (жанды затша) тіл қатқан. Желді ысқырып шақырған, құйынды қуалаған, ауруды көшірген, түрлі жақсылық-жамандықтың иесі бар деп түсініп, оларды бірде шошындырып, бірде тамақ беріп ырза қылып отырған. Әр рудың иесі, қорғаны бар деп есептеп, соларды жәрдемге шақыратын болған. Келе-келе осы ұғым-түсініктен дін туған. Одағайдың қарапта сөзсіз (шақыру сөзінсіз) болмайтыны – сол әртүрлі табиғат күштеріне: ай күнге, малға, т.т. қарап сөйлеп, соларға көңіл айтып, соларды жәрдемге шақырудан қалған қалдық. Кейін адам баласы оны ұмытып, әр одағай өзінің әніне қарай түрліше көңіл райын білдіретін сөз болып кеткен» [91]. «Сондықтан антикалық тілтанушылар тіл мен сөздің сиқырлы-магиялық қасиеттерін жоғары бағалаған. Олар ғарыш, табиғат және адамды бір бүтін ретінде тани отырып, тілді философияның бір бөлігі ретінде қарастырған» [92, 445].

Мифонимдер – аңыз-әңгімелерде, мифтерде сақталған тылсым, қасиетті саналатын, адамдар оларды шынайы бар деп түсінетін атаулар. Теонимдер – құдай, әулие атаулары және шығармалардағы діни атаулар. Осы атауларға белгіленім қою үшін сәйкес кодталған алгоритмдерді құру, қазақ ұлттық корпусына енгізу ономастика, этнолингвистика, мифология, лексикология, лексикография, корпустық лингвистика және т.б. бағыттар үшін ғылыми маңызы бар жұмыс.

Қазіргі таңда ҚТҰК-да белгіленім түрлері бойынша көптеген тілдік таңбалар жинақталды. Соның ішінде, қазақ тіліндегі ұлттың болмысын, тұрмыс-тіршілігін, мәдениеті мен өзіне тән өмір сүру ерекшелігін танытатын 1200 лингвотанымдық бірліктер мен таңбалардың тілдік және мәдени-семантикалық мағынасы ашылды. Олардың белгіленімі (сөз табы, стильдік белгі, мәдени (заттық) белгі қойылды және сөз мазмұнындағы мұндай ақпараттар түсіндірмеде айқындалып, көрнекі материалдар (иллюстрация) арқылы нақтыланып отырды [40].

Сондықтан, жалпы ономастика ғылымы атаулардың тек тілдік мағынасын ашумен, яғни тілдік кодпен шектеліп қалуына мүлде болмайды. Тіл 3 түрлі кодқа салынады. Өйткені ономастикалық атаудың мазмұнында 1) тілдік 2) мәдени 3) тарихи кодтар сияқты сан алуан ақпараттар болады [93]. Себебі тілдің түрлі қабатындағы сан алуан білімдердің тіл иелмендеріне көбіне тек тілдік коды ғана, үстірт ақпарат қана қолжетімді болады. Тілдік код дегеніміз – ақиқат дүние туралы білімді сақтап тұрған тілдік бірліктер, атаулар. Ал мәдени кодтарда ерте заманнан қоршаған орта туралы халықтың өмірлік тәжірибесі мен білімі, дәстүрлі этникалық мәдениеті, наным-сенімі, салты, дүниетанымы көрініс тапқан тілдік емес ақпараттың терең қатпары жасырын тұрады.

Ономастикалық бірліктер ұлттық код болып саналады. Олардың мазмұнында мәдени, тарихи, этностың болмысынан, мәдениетінен, тарихынан мәлімет беретін ақпарат жинақталады. Сондай ақпараттардың бірі – этностың алғашқы дүниетанымы, яғни мифтік, діни дүниетанымы. Әлемнің ономастикалық бейнесі – қоршаған ортаны, әлемді түсінудің ментальды формасы.

Осы орайда ұлттық ономастикалық жүйе халқымыздың, әр заманда мекен еткен халықтар мен ұлыстардың сан ғасырлар бойы жасаған ұжымдық ой-санасының жемісі, ортақ мәдени, рухани туындысы ретінде осы атаулардың бойындағы мәдени кодтар жүйесін ғылыми тұрғыдан қазіргі қазақстандық қоғамдағы қалыптасқан жағдаят контекстінде жаңа сапада жаңғырта терең зерделеуді қажет етеді. Себебі – тіл арқылы ұлт болмысын анықтайтын мәселелердің мазмұны мен түйін-дәйектері ұлттың мәдени кодында астарланған. Ал мәдени кодты ашудың кілті – лингвомәдени таңбаға салынған ақпараттарды аша білу. Ұлттық ономастикалық кеңістігінде шоғырланған ата-бабамыздың бай рухани мұрасы, дәстүрлі рухани мәдениетінің көздері: фольклоры, мифтік, діни, космогониялық танымы, паремиологиялық қоры қазіргі қазақ қоғамындағы кумулятивтік маңызы мен этномәдени ақпаратты тасымалдаушысы ретінде алғаш рет кешенді ғылыми-зерттеу нысаны болып отыр.

Ономастикалық кеңістік немесе ономастикон – шынайы, гипотезалық және таңғажайып объектілерге атау беру үшін халықтардың тілінде қолданылатын жалқы есімдердің саны [94]. Ономастикалық кеңістік жеке секторлардың, аймақтардың жиынтығын құрайды. Атауы бар объектілерге сәйкес оронимия, гидронимия, антропонимия, зооним, космоним т.б. болып көрініс табады.

Қазақ ономастикасындағы онимдік бірліктерді ұлттық корпус арқылы репрезенттелу мәселелері әлі ғылыми тұрғыдан жан-жақты зерделеуді қажет етеді. Бұл, өз кезегінде, сөздің, атаудың мәнін тереңірек түсінуге мүмкіндік береді, тек қазақстандық өскелең ұрпақтың ғана емес, сонымен бірге мемлекеттік

тілді, қазақ халқының тарихын, мәдениетін зерттеуге деген қызығушылықты оятады.

Бұл мәселе қазақстандық лингвистика үшін біршама жаңа болып, қазіргі уақытта жалпы лексиканың (фразеологизмдердің, метрологиялық лексикалардың және т.б.) жеке қабаттарының мәдени-тілдік кодтары үзік түрінде зерттеліп жатыр. Отандық ономастика өзінің жарты ғасырлық дамуында маңызды ғылыми-теориялық, ғылыми-фактологиялық қорды, антропоцентрлік парадигмада болашақ зерттеуге арналған алғышарттарды жинақтады. Қазақ ономастикалық ғылымының бастауында Г.Қонқашпаев [95], Е.Қойшыбаев [96], А.Әбдрахманов [97], Т.Жанұзақов [98], В.Н.Попова [99], Е.Керімбаев [100] және басқа да ғалымдар тұрды. Отандық ономастиканың ғылыми-теориялық негізі кейінгі он жылдарда жобаның салмақты теориялық және ғылыми-методологиялық бастамы болатын лингвокогнитивті бағыттағы Б.Тілеубердиев [84], Г.Мәдиева [101], Қ.Рысбергген [102] және басқа да қазақ ғалымдарының зерттеулерімен толықты.

Қазақ ономастикасының діни-мифологиялық кодын зерттеу, олардың қазіргі қазақ лингвомәдениеті контекстінде семантикасын жүйелеу және жаңғырту мәселесі қойылып отыр.

Ономастикалық атаулардың мәдени-тарихи кодтарының ғылыми-тілтанымдық интерпретациясы рухани-этикалық құндылықтар, халықтық экологиялық таным арқылы дәстүрлі дүниесін тануға, халықты «ішкі жағынан» тануға мүмкіндік береді. Өз кезегінде бұл қазақстандық өскелең ұрпақтың, мемлекеттік тілді үйренуші тұлғалардың қызығушылығын тудыруы тиіс, қазақ халқы туралы, рухани, адамгершілік құндылықтары, ұлттық рухани бағымдылық пен стереотиптері атауларда өрнектелген ақпарат оның дәстүрі, салты білімдердің жаңа көздерін аша алады.

Қазақ тіліндегі ұлттық ономастикалық код жүйесін мифонимдер мен теонимдер ішкорпусын жасау арқылы автоматты лингвистикалық белгіленімдермен қамтылған бағдарламалық әзірлемесін қолданбалы, лингвотеориялық және практикалық негіздері тұрғысынан зерттеуге болады. ҚТҰК-да қазақ халқының мәдениеті мен дүниетанымын көрсететін энциклопедиялық сипаттағы мифонимдер мен теонимдердің ішкорпусын жасау алға қойылып отыр. Ономастикалық ішкорпустағы мифоним, теонимдер жеке сөз, фразеологизм, мақал-мәтел, қанатты сөз, этноним, прецедентті есім, прецедентті мәтін, концепт түріндегі этномаркерлі бірліктермен беріледі.

Әлемнің мифологиялық, теологиялық бейнесінің құрамдас бөліктерін зерттеу қазіргі кезде этнолингвистика, лингвомәдениеттану, лингвокогнитивтік, когнитивті-дискурстық бағыттар аясында қарқынды жүргізіліп келеді. Архаикалық және қазіргі мифологиялық, теологиялық сана ерекшеліктері фольклорлық және көркем дискурстар негізінде қарастырылып келеді. Зерттеу материалдары ретінде лексикографиялық, диалектілік, тарихи дереккөздер, прецедентті мәтіндер мен бір кездері халықтың әлем бейнесінде қалыптасқан атаулар алынып отыр.

Негізгі міндеті ретінде:

– құрамында мәдени-танымдық ақпараты бар мифонимдік, теонимдік бірліктерді жинақтау, олардың мәдени семантикаларын ашу;

– энциклопедиялық сипаттағы репрезентативті лингвомәдени ақпарат беру;

– мифонимдік, теонимдік бірліктер көп қолданылатын мәдени-танымдық мазмұндағы мәтіндер базасын жасау көзделеді.

Қазақ халқының мифологиялық дүниетанымы, діни түсініктері фразеологизмдерде, мақал-мәтелдерде, қанатты сөздерде, фольклорлық мәтіндерде (ертегі, мифтік әңгіме, аңыз, әпсана, жырлар, қисса) көрініс табады. Мысалы, қазақ ұғымы мен болмысында көк сөзінің мағынасы терең. Ол Тәңірдің тұрағы ретінде ұғындырылады. Сондықтан да қазақтар аспанға қарай қолын көтеріп жалбарынады: Тәңір жарылқасын; Тәңір тілеуіңді берсін; көк соққыр; көк бойлады; көкке жетті; көкке көтерді; көк күмбез; көк соққан, т.б. [103].

Жалпы алғанда, түркі тілдерінде сөйлейтін халықтардың бас иетін тәңірісі – «аспан», «көк тәңірі». «Тәңір» сөзі «Құдай, Алла, Жаратушы» деген мағынаны білдіреді. «Тәңірдің жаралуы туралы мифте» Тәңірдің мұхиттан жаратылуы былай деп суреттелінген: «Баяғыда көк те, жер де болмаған тек бір ғана шетсіз-шексіз мұхит болған. Күндердің бір күнінде мұхиттың ортасында ақ жарық көрініп, одан нұр шашқан алтын жұмыртқа шыққан, оның ішінде күллі әлемнің болашақ түп атасы Тәңір жаралған [104]. «Тәңір» туралы фразеологиялық бірліктерде Тәңірден тап; Тәңір атқыр; Тәңірге не жаздым; Тәңір жарылқасын; Тәңір ісі; Тәңір жазса; Тәңір жарылқап [103] сияқты табыну мазмұндас тіркестер де кездеседі.

Мифологиялық шығармалардағы мифоантропонимдердің, соның ішінде демонимдердің бірі – «албасты» ұғымы. Адамзат баласының өзін қоршаған ортада болып жатқан құбылыстарды қабылдауының, алғашқы мифологиялық таным-түсінігінің қайнар көзі мифтер, мифологиялық ертегілер, аңыз-әңгімелер, наным-сенімдер дейтін болсақ, аталған лексикалық бірлікке қатысты мифтерде *албасты* ұзын жалбыраған шашы бар құбыжыққа ұқсас әйел адам ретінде суреттеледі. Ол басқа жануарлар кейпіне немесе затқа ауыса алады, мифке тән бейнеде де айтылады. Г.Потанин өз еңбегінде түркі халқының тілдік қолданысындағы албасты ұғымы жайлы «егер әйел адам босану кезінде есінен танып құласа, онда оны *албасты басты* деп түсінген» деген ойын білдіреді [104, 60]. Ш.Уәлиханов *албасты* сөзіне мынадай анықтама береді: «Албасты – әйелдер босанғанда зиян келтіретін рух. Оларды жезтырнақ деп те атайды. Олардың басшысы сөрел (биік ағаштардың басында мекендейтін мифтік тіршілік иесі), бойы 3 сажын, кеудесі қушық, қалған жерінің бәрі аяқ, тұяғы жіңішке» [105, 182]. Мифологиялық шығармаларда «албастыдай», «албасты әлі буып тұр», «албастыдай болып кеткен», т.б. тіркестер көптеп кездеседі.

«Жалмауыз кемпір» – мифоантропонимі мифологиялық шығармалар тілінде жиі қолданысқа түсетін мифооним. С.Қондыбай өз еңбегінде жалмауыз демонологиясының қартайған кемпір бейнесінде көбіне жеті басты ретінде суреттелетінін айтады. Мысалы, Мұқағалидың өлең жолдарында «Жалмауыздың оқпанындай үңірейген үңгір тұр» [106, 378.], Оралханның шығармасында «Мүмкін, шешесі «жалмауыз» деп сондықтан айтқан шығар, ашынғанда не демеуші еді әйел заты» [107, 98.] деп қолданылады. Мифологиялық және мифопоэтикалық шығармаларды қарастыра келе, «жалмауыз кемпір» онимінің қазіргі кезде семантикалық өрісінің кеңіп, мифтік кейіп-

кер мағынасынан гөрі адамның жағымсыз қасиетін білдіретін ұғымның кең қолданыс алғандығын байқаймыз.

Сонымен қатар, «пері» ұғымы мифологиялық шығармаларда жиі кездеседі. Пері: 1) діни сенім бойынша аспанда, су ішінде тіршілік ететін жамандық пен зұлымдықтың күш иесі; 2) ауыспалы мағынасы – жынды, сотқар, долы кісі; 3) ауыспалы мағынасы – ерекше сұлу, әдемі қыз; 4) атағы шыққан, даңқы жайылған, елді аузына қаратқан адам [108, 360.]. Тілімізде «су перісі», «пері қызы», «перінің елі», т.б. тіркестер жиі кездеседі.

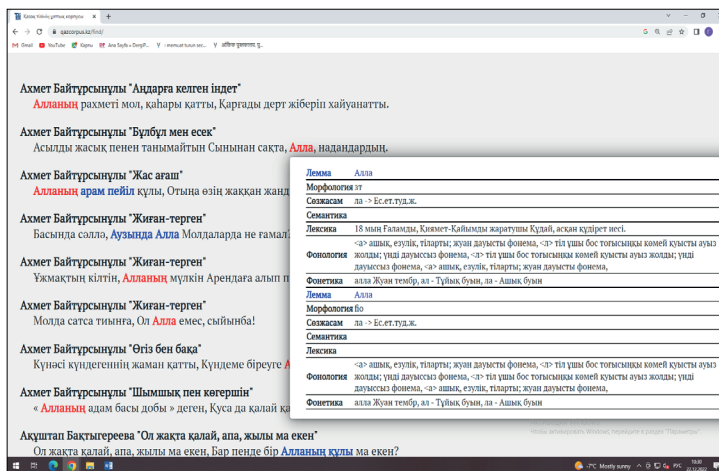
Қарап отырсақ, мифонимдерді мифологиялық шығармалар мен қиял-ғажаып ертегілерден жиі кездестіруге болады. Бұл мәселе төңірегінде онимдерді теріп-жинақтап қана қоймай, олардың мифтік-танымдық сипатын айқындап, соның негізінде фразеологизмдердің қалыптасуы мен қолданысқа ену тарихын зерттеп, ономастиканың жаңа бағытта дамуына жол ашуға болады.

Сөз етіп отырған мифонимдер мен теонимдерді төмендегідей топқа бөлуге болады:

Кесте 5

Мифонимдер мен теонимдер классификациясы

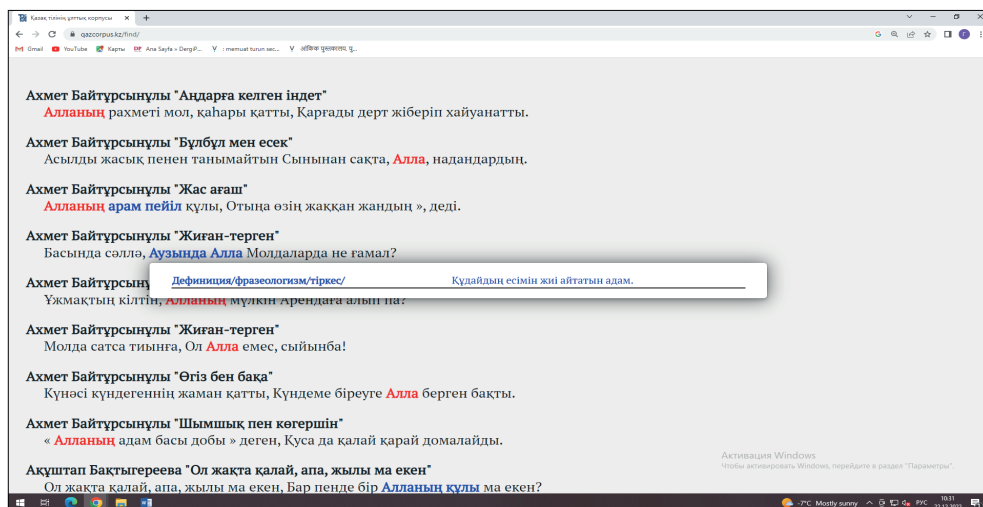
МИФОНИМДЕР		ТЕОНИМДЕР	
Мифозооним	<i>Айдаһар, жезкиік, самұрық, т.б.</i>	Періштелер	<i>Әзірейіл, Жәбірейіл, т.б.</i>
Мифотопоним	<i>Сорқұдық, Арафат тауы, От дариясы, т.б.</i>	Пайғамбарлар	<i>Мұхаммед, Мұса, Иса, т.б.</i>
Мифофитоним	<i>Бәйтерек, дарақ, жекен, т.б.</i>	Сахабалар	<i>Омар, Әбу Бәкір, Осман, Әли т.б.</i>
Мифоантропоним	<i>Әзірейіл, ібіліс, албасты, жалмауыз, мыстан, пері, шайтан, періште, т.б.</i>	Жаратушы	<i>Алла, Тәңірі, Құдай, т.б.</i>



Сурет 10. Алла теонимінің Қазақ тілі ұлттық корпусындағы көрінісі

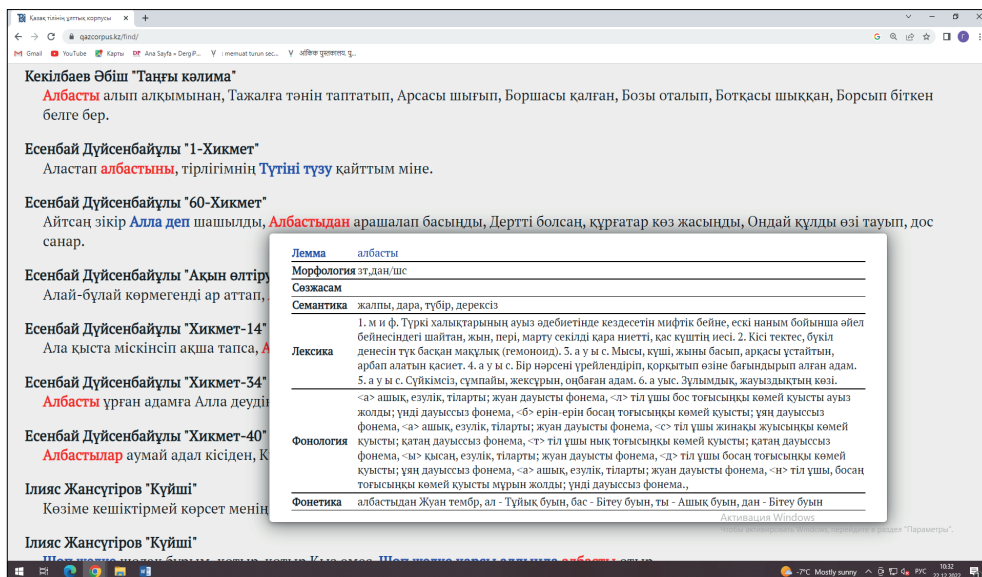
ҚТҰК-да Алла теонимінің тек лексикалық мағынасы ғана берілген. Мәдени-семантикалық мағынасы ашылмаған. Алла теониміне қатысты тұрақты тіркестер де берілген, бірақ толық емес.

Корпусқа теонимдерді белгіленім ретінде толық енгізгеннен кейін жоғарыда аталған мәдени-семантикалық мағынасына қатысты белгіленімі бар ұяшық шығады және тұрақты тіркестер толықтырылады. Сонымен қоса ҚТҰК базасына мақал-мәтелдер, қанатты сөздер, этнографиялық еңбектер қосылады. Алла теонимінің синонимдер қатары қазіргі корпус базасында көрсетілмеген. Теонимдер белгіленімін корпусқа енгізу кезінде аталған және тағы да басқа мифонимдер мен теонимдердің синонимдер қатарын алгоритммен кодтап беру көзделген. Мәселен, Алла сөзінің синонимдік қатарын *Жаратушы, Жаратқан, Тәңір, Құдай, Жаппар ие, Раббы, Тағала* сөздері түзеді. Алланың 99 көркем есімі де белгіленім ретінде қосылып, әр есімі жан-жақты қолданыс аясы мен мағынасына қарай талданып ашылады.



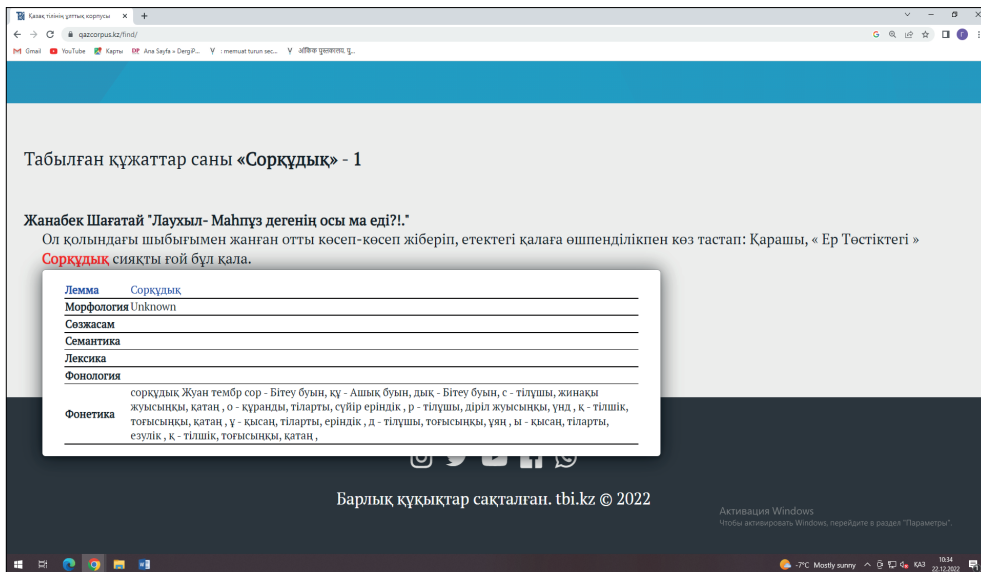
Сурет 11. Алла теонимінің фразеологизмдерде кездесуі

Мәселен, берілген скриншот-сурет корпус базасынан алынған. Бұл жерде тек қана фразеологизмнің дефинициясы берілген, бірақ толық емес. Қазіргі корпус базасында көркем шығармалар, ғылыми мәтіндер, интернет сайттағы газет-журналдар мен әртүрлі ісқағаз мәтіндерінен алынған мысалдар. Содан мифонимдер мен теонимдер ажыратылып көрсетілмеген. Ал мифонимдер мен теонимдер белгіленімі қойылып, өзге мәтіндерден айырмашылығы көрінуі үшін мәтіндік базаны халықтың санасында көрініс тапқан аңыз-әңгіме, ертегі, мифтік әңгімелермен, жырлармен толықтырамыз. Сонда ғана мифонимдер мен теонимдердің жаңадан жинақталған мәтіндері негізінде мәдени-семантикалық мағынасы жан-жақты әрі толық ашылады.



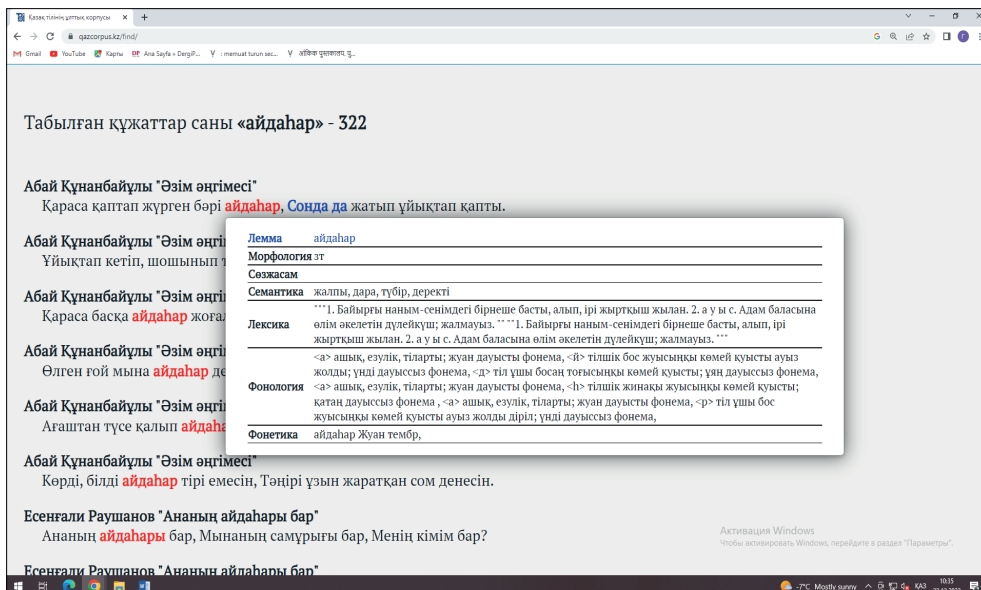
Сурет 12. Албасты мифонимі

Албасты сөзі мифонимдерге оның ішінде мифоантропонимнің демонимдер қатарына яғни зұлым-күш иелеріне кіреді. Скриншот-суретте мифонимге қатысты белгіленім жоқ. Тек қана сөздің мифологиялық шығарма мәтінінен алынғаны немесе мифке жататындығы туралы ғана ақпарат берілген. Корпус базасының албасты сөзіне қатысты лингвистикалық ақпарат беретін бет саны 7 болғанымен, сөздің мағынасын ашуда барлығына бірдей белгіленім қойылған. Аталған сөздің мифоним ба, жоқ оним түрі ме айтылмаған. Сонымен қатар албасты сөзіне қойылған белгіленімдер қатары мен сөз кездесетін корпустағы мәтіндерден фразеологизмдерді кездестіре алмадық. Егер мифонимдер белгіленімі корпусқа енгізілсе, албасты сөзіне қатысты мәдени-семантикалық мағынасы ашылып, корпус базасындағы фразеологиялық қор толығады. Келесі мәселе албасты мифонимінің (оның ішінде демоним) синонимдік қатары көрсетілмеген. Албасты сөзінің синонимдік қатары жын, құбыжық, қырсық, пәле, марту, үббе, құж сияқты демонимдер «албасты басты – қырсық шалды – пәле болды» деп бірін-бірі алмастыра әрі жарыса қолданысқа ие.



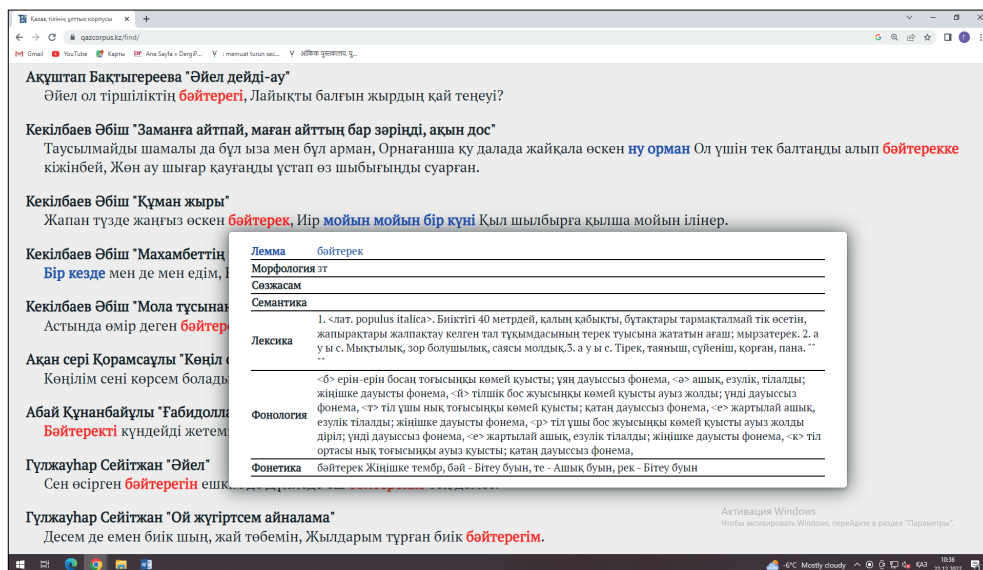
Сурет 13. Сорқұдық мифотопонимі

Мифонимдердің келесі тобы – мифотопонимдер. Мифотопоним бойынша Сорқұдық сөзін іздеп көрпустаң таптық. Сорқұдық сөзінің көрпустаң тек фонетикалық белгіленімі көрсетілген. Ал басқа лингвистикалық ақпарат пен белгіленім түрлері жоқ. Сорқұдық мифотопоним, сонымен қоса сөздердің бірігуі (Сор + құдық) арқылы жасалған сөз.



Сурет 14. Айдаһар мифозоонимі

Мифонимдердің тағы бір тобы – мифозоонимдер. Бұл топтағы атаулар қатарына айдаһар, жезкиік, самұрық, т.б. кіреді. Бірақ корпуста мифозооним деген де, мифтік мәтіндерден алынған атау деген де ақпарат көрсетілмеген. Лингвистикалық белгіленімдерде үстіртін ақпарат беріліп, мәдени коды ашылмаған. Скриншот-суретте алынған айдаһар мифозоонимінің аждаһар, аждаһа, айдағар варианттары көрсетілмеген. Айдаһар мифозоонимімен тұрақты тіркестер де берілмеген.



Сурет 15. Бәйтерек мифофитонимі

Мифонимдердің мифофитонимдер қатарын бәйтерек, дарақ, жекен, т.б. сияқты атаулар түзеді. Скриншот-суретте бәйтерек мифофитониміне қатысты 702 құжат табылып, атауға қатысты белгіленімдері көрсетілген. Бұл мифофитоним бойынша да атаудың мифофитоним екені анықталмаған, тұрақты тіркестер көрсетілмеген.

– ҚТҰК-да жекелеген мифонимдер мен теонимдер белгіленімдері жоқ. Сол себепті лексикалық дефинициясы аздық етеді. Сондықтан корпус базасына ертегі, аңыз-әңгіме, әбсана, мифтік әңгіме, қисса-дастан, жырлар мәтіндерін енгізу қажет. Сонда ғана астарлы, көпшілік біле бермейтін атаулардың семантикасын қысқа-нұсқа, көпшілік халық қабылдайтындай етіп беру арқылы қызық және қолжетімді ақпараттардан тұратын белгіленімдердің мәдени мағынасын, ұлттық, мифтанымдық және діни кодын ашуға;

– Мифонимдер мен теонимдер қатарына кіретін атауларды лингвистикалық талдаулармен беріп қана қоймай, мифтер мен аңыздардан, ертегілер мен жырлардан медиафайлдар, бейнежазбалар мен аудиожазбаларды, суреттерді (кейіпкерлердің, өңірлердің және т.б. суреттері) алып корпус базасына енгізіп, мультимедиялық ішкорпус базасын толықтыра бастауға, оқыту ішкорпусында да, ал ертегілердің аударма мәтіндерін теңестіріп, параллель ішкорпус қорларын толықтыруға;

- Мифонимдер мен теонимдер сөздігін құрастырып тағы бір өнім алуға;
- Мифонимдік, теонимдік атаулардың варианттары мен синонимдер сөздігін түзуге;
- Мифонимдер мен теонимдердің лексикографиялық IT-базасын әзірлеуге болады.

Осылайша, корпустық технологиялар компьютерлік технологиялардың белсенді және қарқынды дамуымен және деректердің үнемі жаңаруымен байланысты заманауи зерттеулердегі әртүрлі мәселелерді шешудің маңызды құралы болып табылады. Нақты әдістерді енгізу арқылы қолданбалы ғылымдар мен гуманитарлық білім салаларына корпустық технологияларды қолданудың айтарлықтай үлесін атап өткен жөн. Әртүрлі бағыттағы зерттеулерге шолу жасағаннан кейін корпус менеджерін, берілетін белгіленім түрлерін өзекті құралдардың бірі ретінде атап өту керек.

Қазақ тіл білімінің ономастика саласы бойынша біраз зерттеулер жүргізіліп, бірқатар диссертациялық жұмыстар қорғалды, атаулар сөздіктері құрастырылды. Алайда, кейбір мифонимдер мен теонимдерге, әсіресе мифтерге негізделген, атаулар бойында сақталып қалған ежелгі дәуірлік этностардың қоршаған әлем бейнесін қабылдаудан туындаған дүниетанымдық көзқарастарының сырын ашу бағытындағы зерттеулерде бірізділік жоқ. Соған орай мақаламызда негіз болған мифтік және діни атауларды көпшілікке қолжетімді интернет желісінде жан-жақты лингвистикалық талданған белгіленімдер арқылы корпуста енгізу технологияларын жинаған материалдарға қатысты салыстыра отырып, осы тектес мифонимдер мен теонимдер атауларына қатысты ортақ пікірлер мен тұжырымдарды бір арнаға тоғыстыруға тырыстық. Сөйтіп, бұл тақырыпта мифонимдер мен теонимдерді ҚТҰК-ға енгізу технологиясы туралы ақпарат бердік. Соның негізінде төмендегі қорытындыны келтіруге болады:

1) тіл үйренушілердің қазақ тілінің мифонимдік және теонимдік атаулары бойынша ұлттық-дүниетанымдық білімдерін тереңдетуге, сол арқылы мемлекеттік тілдің қатынас тілі мәртебесін нығайтуға, отандық және әлемдік мәдени кеңістікте қазақ тілінің мәртебесін, беделін көтеруге септігін тигізеді;

2) мифонимдік және теонимдік мәтіндердің тиімді вербалды құралдарының анықталуы цифрлық кеңістікте мифонимдер мен теонимдерге қатысты ақпараттардың сапалы жасалуына септігін тигізеді;

3) мифонимдер мен теонимдер бар мәтіндердің жазба тілде орфографиялық және пунктуациялық нормаға сәйкестігін, ауызша (аудио, видео) тілде орфоэпиялық нормаға сәйкестігін зерттеу – мифонимдік және теонимдік белгіленімдердің лингвистикалық сауаттылығын арттырады, нәтижесінде мәтіндер стильдік, орфографиялық, пунктуациялық тұрғыдан сауатты құрылатын болады;

4) зерттеу нәтижелері қазақ тілінде жарық көретін қолданбалы бағыттағы оқулықтар, сөздіктер, анықтамалықтар және т.б. құралдардың оңай әрі жедел түрде жарық көруіне ықпал етеді.

5) Жинақталып, мән-мағынасы ашылған мифонимдік және теонимдік тілдік бірліктер, энциклопедиялық сипаттағы лингвистикалық белгіленімдер

арқылы берілген ақпараттар қазіргі антропоөзектік парадигмадағы ономастика, лингвомәдениеттану, этнолингвистика, когнитивтік лингвистика, психоллингвистика, компьютерлік лексикография, т.б. салалар бойынша ізденістердің эмпирикалық базасын кеңейтеді. Нәтижелер қазақ тілінің мифтік, діни кеңістіктерін рухани құндылықтар жүйесі арқылы жаңғыртуға, тілді оқыту үдерісінің тиімділігін арттыруға ықпал етеді.

4 ТАРИХИ ІШКОРПУС

4.1 Тарихи ішкорпус әзірлеудің әлемдік тәжірибесі

Бүгінгі жаһандану кезеңінде өзге елдермен саяси-экономикалық қарым-қатынастарға байланысты ақпарат ағымы күннен күнге көбеюде. Осыған байланысты әлемдік лингвистикада тілдік корпус мәселесі жан-жақты талқыланып, өзекті мәселеге айналып отыр. Қазақ тіл білімінде қолданбалы лингвистика саласы 1970 жылдардан бастау алады. Тіл зерттеуді автоматтандыру мәселесі профессор Қ.Бектаевтың есімімен байланысты [109]. Одан кейін оның шәкірті профессор А.Жұбановтың «Основные принципы формализации содержания казахского текста» монографиясы қолданбалы лингвистиканың дамуына өз үлесін қосты [110]. Соңғы уақытта Ахмет Байтұрсынұлы атындағы Тіл білімі институтында «Қолданбалы лингвистика» бөлімі тілдік корпусы ғылыми-теориялық және практикалық тұрғыдан зерттеу және қазақ мәтіндерінің базасын өңдеп, корпусқа енгізу мәселесімен айналысып келеді. Электронды тілдік корпус халық игілігі үшін, кез келген тілдік ақпаратты пайдаланушы оңай әрі жылдам табу үшін қажет. Қазіргі таңда әлемдік қолданбалы лингвистика саласы бойынша тілдік корпусың барлық функционалды стильдерін қамтитын ішкорпустары әзірленуде. Атап айтқанда, «екі тілді немесе үштілді параллель корпус», «тарихи корпус», «ауызша корпус» т.б. Жалпы аталған басқа ішкорпустарға қарағанда «тарихи ішкорпусты» әзірлеуде аса үлкен жауапкершілікті қажет етеді. Себебі ерте кезеңдегі жазба ескерткіштер мәтіндері әртүрлі графикада және әртүрлі көлемде сақталған. Осыған орай қазіргі таңда «тарихи ішкорпустың» әлемдік тәжірибедегі қолданысын зерттей отырып, қазіргі қоғамға өзінің тарихи мәдени құндылықтарынан мол ақпарат ала алатындай қазақ тілінің тарихи ішкорпусын әзірлеудің маңыздылығы зор. Тарихи ішкорпуста қазіргі қазақ тілі тарихының әртүрлі кезеңдері жазба мұраларынан, тіл тарихынан, сөздің этимологиясынан, т.б. тілтанымдық тұрғыдан ақпарат беретін мәліметтер жиналады. Тарихи мәтіндер базасын жинаудың алғашқы жұмыстары 1928 жылы басталған. Осы жылы ортағасырлық жазба ескерткіш «Кодекс куманикус» мәтіні алғаш реті ЭЕМ жадына енгізілген. Отандық түркітану ғылымы үшін «Куманша-қазақша жиілік сөздігінің» алынуы қазақ тіл білімі үшін аса бағалы зерттеу жұмысы болған [12]. Қазіргі жаңа компьютерлік технологиялар тарихи материалдарды өңдеудің жаңа корпусын қалыптастыруға мүмкіндік береді. А.Фазылжан: «Тарихи субкорпустың болуы шарт, тарихи кезеңдердегі сол тілдің көрінісін білдіретін мәтіндердің белгіленім қойылған базасы болу керек. Мәселен, ол база болса, XVI-XVIII ғасырлардағы қазақ тілінен хабар ала аласыз немесе бір сөздің тарихи кезеңдердегі көрінісін таба аласыз», – деп атап көрсетті [111].

Қазақ тілінің жазу тарихында бірнеше (көне, араб, латын) графика болғаны белгілі. Сол жазулардың ішінде бүгінге дейін араб графикасымен жазылған жазба мұралар қоры мол сақталған. Оның өзін қадим, жадид, төте жазуымен сақталған қолжазбалар деп бөлуге болады. Бүгінгі таңда мәтіндер базасын жаппай цифрландыру кезеңінде араб графикасымен жазылған жаз-

ба мұраларды электронды форматқа айнылдыру аса оңай жұмыс екені баршаға мәлім. Бұл жайында түрік ғалымы Халит Ерен: «...коммуникациялық кеңістікте түркі халықтарына ортақ жазба мұраларымыздың мәтіндерін теріп, енгізудің лингвистикалық программалаудың жолдарын қарастыру қажет», – дейді [111]. Жазба мұралар мәтіндерінің электронды нұсқалары тіл тарихымен айналысатын зерттеушіге немесе қызығушылық танытып жүрген кез келген пайдаланушыға қолжетімсіз болатыны жасырын емес. Олардың көбі шетелде немесе отандық мұражайлар мен кітапханалардың сирек қолжазбалар қорында сақтаулы тұр. Өзге елдердің тарихи ішкорпусын әзірлеудің тәжірибесіне назар аударатын болсақ, мәселен, Ресейде тарихи корпус ертерек басталған. Содан болса керек, орыс тілінің корпусы өзге тілдердің корпусына қарағанда жақсы әзірленген [112]. Орыс тілінің тарихи корпусы іштей көне орыс тілі, қабық хаттар (аққайыңның қабықтарына жазылған көне жазулар), ескі орыс тілі, шіркеулік славян болып бөлінеді. Мәселен, 2007 жылы көне орыс тілінің қабыққа жазылған барлық грамоталары корпус базасында қамтылған. Базада әр грамотаның нөмірі, табылған жері, құжаттың түрі, грамотаның мәтіні, транскрипциясы, аудармасы енгізілген [39].

Ескі орыс тілі ішкорпусы XV-XVIII ғасырлар аралығын қамтитын үш миллион сөзқолданыстан тұрады. Корпусқа әдеби шығармалар, шежірелер, іскери хаттар, мақалалар тізімі, қабық хаттар, тұрмыстық деңгейде жазылған хаттардың мәтіндері енген. Орыс тілінің тарихи корпусының ерекшелігі – пайдаланушы өзіне қажетті ақпаратты тауып қана қоймайды, конкурдастағы кез келген сөзге меңзермен нұсқағанда ұяшықта сол сөз туралы лингвистикалық ақпараттар ала алады. Айталық, «яблоко» сөзін меңзермен басқанда, «яблоко» сөзіне грамматикалық (сущ, вин, с, ед) анықтама беріледі және бұл сөз қанша рет және қанша құжатта кездесетіні туралы толық ақпарат ұсынылады [113].

Орыс тілінің тарихи ішкорпусы жыл сайын жаңа жазба мұралармен толығып, жаңарып тұрады. Мәселен, 2015 жылы 7 мың сөзқолданыстан тұратын XIV-XVII ғғ. жаңа мәтіндер (шежіре, іскери құжаттар) енген болса, 2020 жылы 90 мың сөзқолданыстан тұратын XIII-XVI ғасырлардағы шығыс славяндардың іскери мәтіндері қамтылған. Мәтіндер морфологиялық белгіленімсіз, тек [*] символы арқылы ашылады.

Жалпы әлем тілдерінің корпусын барласақ, көптеген тілдерде тарихи ішкорпусының мәтіндер базасы жасалмаған. Бұны орыс тілінің ұлттық корпусының сайтында берілген тізімінен байқауға болады. Мәселен, сайтта әлем тілдеріне жасалған классификация бойынша орыс тілінің тарихи ішкорпусын есептегенде, герман тілдерінен: американ ағылшын тілінің тарихи корпусы, ескі нидерланд тілінің тарихи корпусы, исланд тілінің тарихи корпусы; роман тілдерінен: орта ғасыр француз тілінің корпусы, үндіеуропа тілдерінен: валлы тілінің тарихи корпусы, үнді-еуропа емес тілдерден: оксфорд көне жапон корпусы, грузин тілінің корпусы, түркі тілдерінен: татар тілінің тарихи корпусы жасалғанын көрсетеді. Сондай-ақ аталмыш тілдердегі онлайн іздеу жүйесі бойынша олардың бірі қолжетімді болса, кейбірі қолжетімсіз болып тұр.

Неміс тілінің ғалымдары, атап айтқанда, проф. Марсель Эрдал, проф. Клаус Рёрборн және проф. Петер исламға дейінгі көне түркі тілдерінің элек-

тронды корпусын әзірлегенін «<https://vatec2.fkidg1.uni-frankfurt.de>» сайтынан анықтадық [113]. 1999-2003 жж. жазба мәтіндер HTML форматында енгізіп және қайта редакцияланғаны туралы ақпарат бар. Корпусқа Алтын Ярлұқ, соғды, көне түркі, ұйғыр жазуымен жазылған мәтіндер, т.б. көне түркі тілдерінде жазылған жазба ескерткіштердің мәтіндері қамтылған. Жазба ескерткіштің метасипаттамасы мен лингвистикалық белгіленімдері жүйелі берілген. Мәселен, Алтын Ярлұқ жазба ескерткішінің мәтінін бөліктерге бөліп, сілтеме арқылы көрсеткен. Мәтінде кездесетін сөздердің транслитерациясын уулқу, транскрипциясын үлкі, морфемасын үлкі, аудармасын Vieh кестеге салып, аудармасын жанына көрсеткен. Бұдан түйетініміз, жазба ескерткіштегі әрбір сөзге кесте арқылы берілген талдаулары кез келген пайдаланушыға өте бағалы құрал болатыны аңғарылады.

Қазақ тілінің тарихи мәтіндер базасын төмендегі кезеңдер бойынша, яғни тарихи ішкорпусқа топтастыруға болады:

1. V-IX ғғ. көне түркі дәуіріндегі жазба ескерткіштерінің мәтіндік базасы;
2. X-XV ғғ. орта ғасыр дәуіріндегі жазба мұралардың мәтіндік базасы;
3. XVI-XIX ғғ. ескі қазақ жазба үлгілерінің мәтіндік базасы;
4. XX ғасырдың басындағы жазба мұралардың мәтіндік базасы;
5. Латын графикасымен жазылған жазба мұралардың мәтіндік базасы.

Тарихи ішкорпусқа енгізілетін аталмыш материалдардың көбі араб графикасымен оның ішінде қадим, жадид, төте жазуларымен жазылған. Бұндай түрлі графикада сақталған жазба ескерткіштер мәтіндеріне қызығушылық танытатындар – зерттеушілер (түркітанушылар, тарихшылар, әдебиетшілер).

Араб графикасымен жазылған жазба ескерткіштер мәтіндерін корпусқа енгізу мәселесін қарастырайық:

- корпусқа әр дәуірдегі көлемі әртүрлі жазба ескерткіштер алынады;
- әрбір жазба мұраға ақпараттық белгіленімдер беріледі;
- мәтіндегі әрбір сөзге лингвистикалық белгіленімдер қойылады.

Қазақ тілінің тарихи корпусы алғаш рет жасалып жатқандықтан, XII-XX ғасырлардағы әр дәуірден екі-үш жазба мұрадан ғана алынады. Әзірше араб графикасымен жазылған қолымызда бар жазба мұралармен жұмыс жасалады. Материалдарды тезірек жинақтау үшін, мәтіндердің дайын электрондық нұсқалары ашық электронды кітапханалардан алыну керек. Әрине, араб графикасымен сақталған материалдардың, әсіресе жарық көрмеген жазба мұралардың көбін сирек қолжазбалар қоры сақталатын кітапханалардан табуға болады.

Қазақ тілінің тарихи мәтіндер базасы әртүрлі жанрдан тұратыны белгілі. Сондықтан базаға шежіре, дастан, жыр-өлең, проза, оқу құралы, мақала, хат, жарлықтар, т.б. жанр түрлері енеді.

Әр жазба мұраның факсимелесі титулынан бастап барлық беті сканерленіп, жеке-жеке бет ретінде көрсетіледі. Сканерленген мәтіндердің көрінісі көзшалымға ыңғайлы болу жағы назарға алыну керек, себебі көмескі тартқан кейбір жазулардың сапасы оқуда қиындық тудыруы мүмкін. Сонымен қатар араб графикасымен жазылған әр беттің кирил графикасындағы транскрипциясы мен аудармасы беріледі. Осы уақытқа дейін әлемдік және отандық

түркітанушы ғалымдар араб графикасымен жазылған қолжазба мәтіндерін ба-рынша өз тілдеріне жақын және әртүрлі графикамен транскрипциялап келді. Қазіргі зерттеушілер бірқатары жазба ескерткіштер мәтіндерімен жұмыс жа-сағанда дайын транскрипцияланған мәтіндерді пайдаланады. Корпусқа жазба мұралардың түпнұсқасымен берілетін мәтіндер ізденушінің мәтіндердің шы-найы қалпын көруге, талдауына мүмкіндік беретіні сөзсіз. Бұл үшін кез келген пайдаланушыға қолжетімді болатындай жазба мұралардың факсимелелерін жинақтап, сұрыптап, сканерлеп, корпусқа енгізуді қажет етеді. Араб графика-сымен жазылған мәтіндерде кейбір бұлыңғыр жазылған әріптер, сөздер, тіпті сөйлемдер жиі кездеседі, сонымен қатар мағынасы көмескі келетін сөздердің мағынасын анықтауда қиындық келтіреді. Бұл ретте ізденушіге оңай әрі ыңғайлы болу үшін берілген жеке беттегі мәтіннің аудармасы қоса беріледі. Демонстрациясы қалай болады дегенде, <http://gramoty.ru/> «Древнерусские бе-рестяные грамоты» сайтындағы алгоритімін ескеруге болады [114]. Яғни, мәтіннің факсимелесі, транскрипциясы, аудармасы.

Әрбір жазба мұраға берілетін ақпараттық белгіленімдері толық, дәл болуы керек. Әлемдік тарихи корпусның тәжірибелерін ескере отырып, әрбір жазба мұраға берілетін ақпараттық белгіленім төмендегідей болуды жоспарлап отыр-мыз:

- мәтін туралы қысқаша аңдатпа;
- мәтін авторы (аты-жөні);
- мәтін тақырыбы;
- мәтін графикасы: көне жазу, араб-қадим, араб-жадид, араб-төте; мәтіннің алғашқы жарияланымы;
- мәтін стилі: ресми, публистикалық, тарихи шежіре, көркем, ғылыми, педагогикалық дискурс;
- мәтін жанры: проза, поэзия, (аралас: поэзия, проза), сөздік, хат, жарлық, газет, өтініш, арыз, оқулық, мақала, ғылыми ақпараттық, ғылыми көпшілік;
- мәтін дереккөзі;
- мәтіннің нұсқалары: Қазан нұсқасы, Демезон нұсқасы,
- мәтін транскрипциясының авторы;
- мәтін аудармасының авторы;
- мәтіннің бет саны;
- аудитория жасы;
- мәтінді корпусқа енгізген автор;
- мәтінді корпусқа енгізген мерзімі;
- Ескерту: жоғарыдан төмен берілген сандар сөйлемнің қатарын көрсетеді.

Сонымен, ғылым мен техниканың үлкен қарқынмен даму кезеңінде жаңа техникалық мүмкіндіктерді пайдалана отырып, тарихи мәтіндердің ішкорпу-сын құрастыру – кезек күттірмейтін мәселе. Тарихи ішкорпус әзірлеудің басты артықшылықтары мынада:

- лингвистикалық белгіленімдер зерттеушілер үшін тек ақпараттық құрал қызметін ғана емес, өздеріне қажетті анықтамалық дереккөз де бола алады. Зерттеуші кез келген жерде және кез келген уақытта отырып, бастаған зерттеу жұмысын жалғастыра алады;

- факсимеленің әр бетіне сәйкес берілетін мәтіннің транскрипциясы мен аудармасы зерттеушінің тарихи-салыстырмалы, диахронды-сихронды зерттеулер жасауына мүмкіндік береді;

- нұсқалар арасында лингвотекстологиялық зерттеулер, лексикографиялық жұмыстар жасауға болады. Бұл әрі қарай архаикалық тілдік (фонетикалық, морфологиялық, лексикалық, фразеологиялық), т.б. белгіленімдерінің модельдерін нақтылауға көмектеседі.

Қазақ тілінің тарихи ішкорпусын әзірлеудің мәселелерін қарастырғанымызда, мынадай нәтижелерге қол жеткізілді:

- біріншіден, қазақ тілінің тарихи ішкорпусын әзірлеудің әлемдік тәжірибесі ескерілді;

- екіншіден, қазақ тілінің тарихи мәтіндер базасын көне түркі, орта түркі, ескі қазақ жазба үлгілері, XX ғасырдың басындағы жазба мұралар деп 4 кезеңге топтастыруға болады;

- үшіншіден, пайдаланушыға ыңғайлы болу үшін сканерленетін жазба мұраның факсимелесінің әр беті жеке-жеке беріліп, сол мәтіннің транскрипциясы мен аудармасы демонстрацияланады;

- төртіншіден, V-XX ғғ. жазба мұралардың стилі, жанры, т.б. ерекшеліктеріне қарай ақпараттық белгіленімдердің тізімі ұсынылады.

4.2 Орыс тілінің тарихи корпусын жасау тәжірибесі мен түркілік лексикалық қабаттың берілуі

Қазақ тіл білімін цифрландырудың алғышарты ретінде тарихи корпусы іске қосу арқылы корпусық лингвистиканың жаңа бағытын дамыта аламыз. Ф.ғ.д., профессор Асқар Жұбанов: «Автоматтанған қазақ тіліндегі мәтіндер корпусының компьютерлік базасын құру ғылыми және практикалық жағынан аса құнды мәселе» екендігін айтады [12]. Ғалым А.Е. Кибриктің ойынша, корпусы жасау микроскоп, телескоп, томография, рентген аппараты және т.б. ойлап табумен салыстыруға болады. Осыдан-ақ корпусың мән-маңызын түсінуге болады.

Отандық ф.ғ.д., проф. А.Қ. Жұбанов, Э.Д. Сүлейменова, А.Жаңабекова, А.Фазылжан, М.Нұрланның, Ресейлік А.Е. Кибрик, В.А. Плунгян, Н.В. Перцов т.б. ғылыми еңбектері қарастырылып, басшылыққа алынды.

Профессор А.Жаңабекованың пікірінше: «Қазақ тілінің ұлттық корпусы – қазақ тілі бойынша барлық мәліметтерді жинақтаушы банк қана емес, ол «ақылды», мамандандырылған білім қоры, үнемі толығып отыратын жүйе ретіндегі мемлекеттік тілдің ақпараттық-инновациялық базасы. Ғылыми мәтін даярлауда пайдасы зор. Лингвистикалық ереже шығару үшін мыңдаған мысалдар, дәлелдер қажет. Нақты мысалдардың негізінде ереже жасалады. Ал миллиондаған мысалдар корпуста бар. Корпус – мәтіндер қоры» [115].

Метабелгіленім – мәтіндердің атауы, авторы, жылы, стилі, жанры, формасы, аудиториясы, шыққан жылы, көлемі, сөзқолданыс саны т.б. бойынша сипаттайтын компьютерлік бағдарлама.

Ғалым В.А. Плунгянның айтуынша, «Корпус тек тілді зерттеудің күшті қаруы ғана емес, сондай-ақ жаңа идеологиясы». «Корпуста бар дүние» мен «тіл-

де бар мәлімет» екеуі бір ұғым емес, екіншісінің ауқымы үлкен болғандықтан біріншісінде көрініс таппай қалуы мүмкін [116].

Тіл деңгейлеріне қарай бірнеше бағыттарға сараланады: лексика-фразеология (дефиниция, сөздік), семантика (полисемия, омонимия, лексика-семантикалық топтар), фонетика-фонология (дыбыс, буын, орфоэпия, просодика), морфология (түбір, қосымша, граммема), сөзжасам (туынды сөз, сөзжасамдық тәсіл, күрделі сөз), синтаксис (тіркес, сөйлем мүшесі, сөйлем түрлері, модельдері), стилистика (стильдік белгілер, стиль, жанр), ономастика (мәтіндегі жалқы есімдер туралы ақпараттар, оларды белгілеу, ономастикалық база), терминология (ғылыми стиль, терминологиялық мағына, терминологиялық база), тіл тарихы (тарихи корпус, тарихи мәтіндер, тарихи белгіленімдер, тарихи сөздіктер), этнолингвистика (ұлттық-танымдық, этномәдени ақпараттар).

Корпустық лингвистиканың бір тармағы – тарихи корпус. Тарихи корпустық лингвистика – компьютерлік технологияларды қолдана отырып, тарихи мәтіндер корпусын жасау мен оны пайдаланудың жалпы принциптерін зерттейтін қолданбалы тіл білімінің жаңа бір саласы. Тіл тарихының корпусын құрау үшін ескі қолжазба мәтіндерінің жиынтығы қажет. Тарихи мәтіндер қолжазбалар мен олардың транскрипциялары, транслитерациялары, аудармалары, тарихи картотека қоры, тарихи белгіленімдер, қажетті ақпараттар мен дереккөздерден құралады. Тарихи белгіленімдердің ара-жігін ашып, тарихи фразеология, тарихи терминология, тарихи морфологиялық деген секілді белгіленімдерін ажыратып көрсете білу корпустық лингвистиканың алдында тұрған зор міндеттердің бірі.

Компьютерлік тілдік қор арқылы қазақ тілінің тарихын жан-жақты зерттеуге жол ашылады. Зерттеушілер ауқымды материалдармен электронды түрде танысып, материалдарды санаулы секундтарда тез тауып алуға, оларға тиісті өңдеулер жүргізуге, тарихи мәтіндер бойынша электронды түпнұсқасы мен кирил эрпіндегі аудармасы мен транскрипциясын, мәтін туралы қысқаша аңдатпа, мәтін графикасы, жазылған жылы/ғасыры, басылымы, мәтін авторы, жанры, сақталған орны мен нұсқаның түрі, мәтіннің бет саны, мәтінді корпуста енгізген маман туралы т.б. ақпарат ала алады, сонымен қатар зерттеудің шынайлығына, нақты әрі дәлелді болуына жол ашады. Тарихи корпус зерттеушілерге қазақ тілінің тарихы бойынша ғылыми зерттеулерді оңайлату және тездету, тиімді ету үшін қажет.

Қазақ тілінің тарихи мәтіндер базасын тарихи ішкорпуста қалайша топтастыруға болатындығы осы саланың маманы ғалым А.Сейтбекованың зерттеу мақаласында кеңірек жазылған [117].

Тарихи картотека қоры тарихи корпустың қалыптасуына ықпал етті. Корпустық лингвистиканы өмірге әкелуге себепші болған бағыт: «картотекадан корпуста» немесе қағаз картотекадан электронды картотекаға айналуы. Корпус – дәстүрлі картотека қоры мен қолдану әрекетінің заманауи, цифрлық, заңды жалғасы.

Ахмет Байтұрсынұлы атындағы Тіл білімі Институтында 5 млн. карточканы сканерлеп, корпустың сайтына пайдалану, реттеу т.б. жұмыстарына «Қолданбалы лингвистика» бөлімінің мамандары атсалысып келеді [118].

Тарихи корпус саласын қалыптастыру, дамыту арқылы біз грамматикалық теорияларды жаңаша зерттеп, жаңа, дәйекті білімдерге қол жеткізіп, тілдік құбылыстарды синхронды және диахронды түрде терең зерттеуге, лингвистикалық зерттеулердің эмпирикалық базасын кеңейтуге, тілдің және оның салаларының дамуы мен тілдік өзгерістер туралы мол ақпарат алуға, лексикалық, грамматикалық және стилистикалық тілдік құралдардың қолданысына статистикалық мониторинг жүргізуге мүмкіндік аламыз. Мысалы, сөздердің қолданысы арқылы олардың қалыптасуы мен мағыналық дамуын анықтаймыз.

Қазақ тілінің тарихи корпусы тарихи бірліктер жайлы анықтамалар мен статистикалық мәліметтер алу мақсатында пайдаланылады. Тарихи корпус-ты қолдану барысында: тарихи сөзформалар, тарихи лексемалар, тарихи грамматикалық категориялардың қолдану жиілігі жайында деректер алуға; белгілі бір кезеңдегі, ғасырдағы тілдік деректердің өзгеріске ұшырау динамикасын бақылап отыруға және оны зерттеуге; әртүрлі жанрдың және әр автор шығармалары тілінің тарихи лексика-грамматикалық сипатына талдау жүргізуге; көпәспектілі лексикографиялық жұмыстар жүргізуде дереккөз ретінде де, мысалы әртүрлі тарихи және заманауи сөздіктерді дайындауда да зор қызмет атқарады; корпус деректерін грамматикаларды құрастыру мен нақтылау және тілді оқыту мақсатында пайдаланылады. Тарихи корпус-ты субкорпустарға қарай әр стиль мен әртүрлі жанрға байланысты саралайтын болсақ, ресми стильден – хан жарлықтары, публицистикалық стильден – «Дала уалаяты» мен «Түркістан уалаяты», «Қазақ» т.б. ескі газеттер, көркем әдебиет стилінен – халық ауыз әдебиеті үлгілері, Қожа Ахмет Яссауидің хикметтері мен Әбілғазы Баһадүрдің қолжазбаларын, ғылыми стильден – XIX ғ. аяғында шыққан сөздіктер мен XX ғ. басында жарияланған оқулықтарды, мәселен Әлихан Бөкейханның «Дүниенің құрылысы» (1926 ж.) астрономиялық оқулығы т.б. араб және бүгінгі күнгі кирил әліпбиімен беру алда тұрған үлкен міндеттердің бірі.

Тарихи корпус-тың практикалық маңызына тоқталар болсақ, жиілік сөздіктерді; мектеп оқушыларына арналған тарихи мәтіндерді түсіндіруді мақсат ететін глоссарийлер мен шағын сөздіктерді, қазіргі заманғы түсіндірме сөздіктер мен тарихи сөздіктерді, диалектологиялық атластар мен сөздіктерді, тарихи ономастикалық сөздіктерді, тарихи грамматикалық, морфемдік, сөзжасамдық және т.б. сөздіктерді, тарихи тіркесімділік сөздігін, антонимдік, синонимдік, омонимдік, паронимиялық және т.б. сөздіктерді, тарихи стильдік сөздіктерді (іскери тіл, публицистикалық, ғылыми мәтіндердің және т.б.), тарихи орфографиялық сөздіктерді құрастыруға септігін тигізеді. Тарихи корпус-ты жасау – қазақ тілінің тарихын тануға, зерттеуге бағытталған жаңа сала, жаңа деңгей. Тіл тарихи бойынша зерттеулерді «корпусқа дейінгі және кейінгі» деп қарастыруға болады.

Тарихи корпус – қазақ тілінің ескі кезеңдерден бері жарық көрген әртүрлі графикалық жүйедегі тарихи ескерткіштер мәтіндерінің лингвистикалық сипаттамасы бар электронды жиынтық. Сөздердің даму тарихы туралы кең ақымды ақпарат алуға болатын платформа.

Тарихи корпусқа тарихи ономастиканы енгізу қажет деп санаймыз. Тарихи ономастикалық базаның құрамы кісі есімдері, ескі қала, өзен аттарымен толығады. Халықтың тарихи жадысының оянуына мүмкіндік туғызады. Тіл жүйесін мұқият қарауға, зерттеуге, жан-жақты білуге мүмкіндік туады.

Тарихи корпустың нысанына тарихи белгіленімдер, тарихи мәтіндер, тарихи сөздіктер енеді. Жазба мұраларды цифрландыру дегеніміз жазуымызды жаңа деңгейде, жаңа форматта сақтау деген сөз. Корпуста сөздіктер мен грамматикалық белгіленімдер, анықтамалық құралдар қамтылады.

Армян тілінің ұлттық корпусы немесе Шығыс армян ұлттық корпусы – Армения Республикасының Мемлекеттік тілі болып саналатын шығыс армян тілінің әдеби нұсқасындағы белгіленген мәтіндердің кең коллекциясы негізінде анықтамалық-ақпараттық жүйе болып табылады. Армян тілінің ұлттық корпусына 19 ғасырдан бастап бүгінгі күнге дейінгі қазіргі Шығыс армян тарихының әртүрлі кезеңдерін қамтитын диахрондық бағдарланған корпус енген.

Шығыс армян ұлттық корпусы (ЕАМС) – Армения Республикасының Ұлттық кітапханасымен бірге (ВАНК) ХІХ ғасырдың 70-ші жылдарынан бастап шамамен 1990 жылдарға дейінгі армян мерзімді басылымдарының үлкен үлгісін фотокошіру және дигитализациялау жобасын жүзеге асырды. Архивте 60 мерзімді басылымның 2668 нөмірлерінен жинақталған жалпы көлемі шамамен 12,7 млн. сөз қолданыс қамтылған. Мысал ретінде, ХІХ ғасырдың аяғы – 1875 жылғы 7 тамыз аралығында шыққан ең ықпалды армян мерзімді басылымдарының бірі «Мшак» күнделікті газетінің бір шығарылымының төрт бетінің фотокошірмесі армян тілінің өзінің ұлттық графикасымен берілген.

Орыс тілінің ұлттық корпусының (<https://ruscorpora.ru/new/>) ішіндегі тарихи корпусын қарағанымызда тілдік деректердің бай әрі жан-жақты қамтылғанын байқадық. Бүгінгі күні орыс тілінің жалпы корпусында 250 миллион, ал тарихи корпусында 13 млн. сөз қолданыс бар, бұл күн сайын толығып отыратын үдеріс. Диахрондық мәліметтер жиынтығы үш кезеңді қамтиды – 1700-1916, 1918-1991 және 1992-2016 жылдар. Қоғамның дамуына байланысты шамамен үш тарихи кезеңіне «Кеңестік дәуірге дейінгі», «Кеңестік» және «Кеңестік дәуірден кейінгі» уақытқа сәйкес келеді. Осы кезеңдердің әрқайсысы UTF-8 кодталған үлкен мәтіндік файлмен ұсынылған, онда бастапқы сөйлемдер кездейсоқ ретпен орналасқан. Мәтіндердің тұтастығын бұзу авторлық құқықты қорғау талаптарына байланысты реттеледі.

Ескі орыс тілінің корпусында (старорусский корпус) 7 493 құжат, 8 561 261 сөз қамтылған.

Көне заманнан бері қолданылып келе жатқан «книга» сөзін терсеңіз, бұл сөз бойынша 77 құжат, 612 сөз қолданыс шығады. 1470 жылдардың (бұдан да бұрынғы жылдардағы қолданысы бар) «Тайная Тайных» деген туындысынан бастап бүгінге дейін қолданысын, яғни ХV-ХХІ ғғ. аралығындағы дереккөздерін, әр сөзге берілген грамматикалық белгіленімдерін көре аласыз. (<https://processing.ruscorpora.ru>) Ал егерде «книга» сөзін негізгі корпуста іздесе 7 437 құжат, 21 510 сөзқолданыс шығады.

Назар аударатын тұс – ескі орыс тіліндегі сөздердің қолданыс жиілігі графеммамен де, табицамен де берілуі. Бұл орыс тілінің тарихи корпусының бір ұтымды тұсы деп білеміз. (<https://processing.ruscorpora.ru/>)

Мәселен, «книга» сөзінің 1395-1730 жылдар аралығындағы қолданыс жиілігі көрсетілген. «Книга» сөзі ескі орыс тілінің корпусында ең алғаш рет 1400 жылы «Псковская 3-я летопись: Архивский 2-й список» (1400-1650) атты ресми мәтінде қолданылғанын тарихи корпус арқылы тез таба аламыз:

«Князь Дмитрей Иванович Углецкой да князь Федор Борисович Вологоцкой ходишах Казани; и тогда московских людеи на Волги у Казани потопло много, мая вь 22 день. Псковское взятие, како взять его князь великий Василен Ивано-вич. От начала убо Руския земли сеи убо град Псковъ ни коим же князем владом бѣ, но на своей воли живяху в немъ сущии людие. Прежня же убо удельная княжения взять князь великий московский под свою область не во едино время ратию, но по разну, яко же лѣтописная книга пишеть».

Бұл мәтін үзіндісінде ескі дәуірдің тілдік қолданысы (лексика, грамматикалық), графикасы, қолданыс жиілігі және бүгінгі күнгі тілдегі графикадағы аудармасы көрініс табады.

Сонымен қатар тарихи корпус тарихи жиілік сөздікке нақты мәлімет бере алады. Мәселен «Книга» сөзі 1400 жылы 1 рет, әртүрлі жылдары түрлі жиілікте қолданылып, ең көп қолданылған тұсы 1630 ж. (121 рет) және 1692 (131 рет сөз қолданысы) жылдар аралығына келеді, ал ең аз қолданылған 1636 жылы 1 рет. Бұл жағдай осы жылдары Еуропада баспа ісінің қарқын алып, кітап шығаруда ілгерушілік болды деген ой қорытуға негіз болады.

Ескі орыс тілінде «Руския» қазіргі тілде «Россия, русский», ескіше «ходиша» қазіргі «ходить» сөзі екенін зерттеуші бір секундта біліп алады. <https://processing.ruscorpora.ru/>

Тарихи морфология бойынша да ақпарат ала алады екенбіз. Ескі орысша «х Казани», қазіргі орыс тілінде «к Казани», қазақ тіліндегі «Қазанға» дегендегі барыс септігінің -ға жалғауымен ұқсастықтың барын аңғарамыз. Бұны кезінде О.Сүлейменов «Язык письма» еңбегінде айтқан. Тарихи корпус арқылы зерттеушінің тұжырымдарына тез әрі нақты дәлел табылды. Тарихи корпусның бір артықшылығы туыс және туыс емес тілдердің арасындағы байланысты көрсетуге мол мүмкіндік береді.

Орыс тілінің тарихи корпусынан орыс тіліндегі түркізмдер болып саналатын қазақтың «қазына» мағынасындағы «казна» сөзін іздестірдік. «Казна» сөзі ескі орыс тілінің корпусында (Старорусский корпус) 66 құжатта, 111 рет сөз қолданысы мәтінімен, авторы, жылдарымен көрсетілген. 1395-1775 жж. аралығындағы қолданыстары түпнұсқа мәтіндері мен графикасы, қолданыс саны мен жиілігі көрсетілген. Қолданылу жиілігі тұрғысынан бұл сөздің ең көп қолданылған жылдары 1728-1767 жж. аралығына келеді екен (1-сурет). Себебі бұл экономика саласының сөзі, көрсетілген кезеңдерде орыстар мен түркі халықтарымен, соның ішінде қазақтармен саяси, экономикалық байланысының күшейгенін тарихтан білеміз. Осы тығыз байланыстың тілдік куәсі ретінде «қазына/казна» сөзінің қолданылу жиілігі қарқын алып, әуелі жазба мұраларда, кейін тарихи корпуста көрініс тапты.

Назарымызға іліккен тағы бір сөз – табун. «Табын» сөзінің орысша нұсқасы. Ескі орыс тілінің корпусында 7 құжат, 9 сөз қолданысы тіркелген. Уикипедияда XVII ғасырда орыс тіліне енген деп көрсетілген. Бұның дәлелін

ескі орыс тілінің тарихи корпусының 1685 жылғы жазбасынан кездестіреміз. Алғашқы жазбаға тіркелген мәтін:

Буди, государь, тебе ведомо; табун под Кожухов пригнали, а торг указано дат(ь) августа в 4-м числе. [Грамотка приказчика Г. Щербачева в с. Спасское (1685)].

Бұл тек жазбаға тіркелгені, хатқа түскені ғана, бүгінгі күнге дейін сақталғандары. Дер уақытында жазбаға түспей немесе жазба мұралардың сақталмай қалатын жағдайлары жиі орын алады. Бұған қарап «табын» сөзі «17 ғасырда орыс тіліне енген екен ғой» деген біржақты түсінік қалыптаспауы керек. Екі халық бұдан да бұрын араласқаны тарихтан мәлім.

Аталған сөз 12 ғасырдың мына жазбасынан да көрініс тапқан әрі грамматикалық метбелгілінімі де берілген:

«От Станигостя к Олуче. Вот ты конфисковал у меня кобылу. А я в этом [деле] не виноват. А тебе доносил про табун Жирнежко так оштрафуй [его] на три гривны. [Берестяная грамота 959 (1075-1100)]».

Сундук (сандық) сөзін орыс тілінің тарихи корпусынан қарағанымызда бірнеше графикада берілгеніне куә болдық: 1) XII ғ. қолданылған ескі орыс графикасы (Берестяная грамота 429 (1160-1180)); 2) бүгінгі орыс жазуындағы аудармасы; 3) басқа тілге, ағылшын тіліндегі аудармасы. Бұл аударма ішкорпусының аясына енеді; 4) «сундук» сөзіне грамматикалық сипаттама берілген.

ory	<ul style="list-style-type: none">• монисто оусьрази кожоуха •в-(•) съ пьльд... три отоцька польпьяна и съ цьльцьмо -од... шьсть гоубиць пьрни и възогьловіе лоудиц... ѿ• роужьн-хо а три бѣла оков(ьц)ь мѣд[ан]... гьлько ларь [Берестяная грамота 429 (1160-1180)]
rus	<ul style="list-style-type: none">• Монисто, серьги, две шубы с чехлами, [каких-то женских головных уборов] три с обшивкой, украшенной лентами (или: из лент, разноцветной), и с очельем ..., шесть губок, перины (?) и изголовье, плащи (или: плащ), ... пять крашенных в красное и три белых, окованный ларчик медный, ..., кувшин (рукомойник), сундук. [Берестяная грамота 429 (1160-1180)]
eng	<ul style="list-style-type: none">• A necklace, earrings, two fur coats &, three [some kind of womans head ornament] with trimming, decorated in ribbons (or: made out of ribbons of various colors) and with a headband, six sponges, a down bedspread (?) and a pillow, coats (or: a coat), & five colored in red and three white, a copper-bonded case, a pitcher (washbasin), a chest. [Берестяная грамота 429 (1160-1180). Переводчик R. Kovalev]

Бұл жайтты қазақ тілінің тарихи корпусын жасауда қажетті белгіленім деп санаймыз.

Тағы бір ұтымды тұсы – корпустағы мәтіндерді көшіріп алуға болатындығы. Халықтың өз мұрасы өзіне қолжетімді. Қолжазба бір кітапхананың немесе Институттың, баспаның ақшаға сататын жекеменшік мүлкі ретінде емес, қолжетімді, тегін, жеңіл іздеу функцияларының болғаны жаңа зерттеулердің қарқын алуына негіз болады. Тарихи корпусты жасау – кез келген тілдің жаһандануға жұтылып кетпеуінің бір тетігі деп білеміз.

Орыс тілінің діни мәтіндерінің корпусын (церковнославянский корпус) зерттегенімізде біраз жайттарға қанықтық. Қазақ тіліндегі тарихи теология саласын дамыту, тілдегі діни картинаны көру мақсатында ескі және жаңа діни мәтіндердің қоры жасақталуының қажеттігін аңғардық. Тәңіршілдіктен бастап, әулие-әнбие, пайғамбарлар, қиссалар, Баба түкті шашты әзіз бен ислам мәтіндерін корпуста енгізсе, тарихи теология корпусын дамытуға жаңа қадам басталар еді.

Назар аударатын тұс – ескі орыс тілі орфографиясының корпуста көрініс табуы. Тарихи корпус болған соң бір мағыналы сөздің әртүрлі ғасырда бірнеше формада жазылуы кездесіп отырады. Орыс тілінің діни мәтіндерінің корпусында 3 түрлі орфографиялық белгіленім бар «точная, упрощенная, модернизированная».

Бүгінгі орыс тілінің графикасымен жазылған «Бог» сөзін (<https://ruscorpora.ru/new/search-orthlib.html>) «точная» белгіленім арқылы іздегенімізде ешқандай сөз қолданыс көрсетілмеді, демек, дәл осылай жазылым бойынша тарихи корпуста сөзқолданыс жоқ болғаны. «Упрощенная» белгіленімде «Бог» жазылымымен бір-ақ қолданыс шықты. «Модernизированная» арқылы 40 құжат, 67 сөзқолданыстағы «бóгъ» сөзінің дәйексөздері мен олардың дереккөздері берілді. Бұл біздегі «Алла, Аллах» деп әртүрлі жазылып жүрген сөздерді жазуда түрлі орфографиялық белгіленімдерді қолдануға болатындығын көрсетеді. Бұл өз кезегінде тарихи орфографияның дамуына негіз болады. Бұл мәселеден қазақ тілінің тарихи орфографиясының жаңаша зерттелуіне тарихи корпустың берері зор екендігін бағамдаймыз. Қазақ тілінің тарихи фразеология корпусын бастауды да күн тәртібіне қою керек деп санаймыз.

Қырым татарларының корпусы, «Туған тіл» атты татар тілінің ұлттық корпусы, татар тілінің жазбаша корпусы, башқұрт тілінің поэтикалық корпусы, түрік тілінің корпусы, түрік тілінің ауызша корпустарын қарастырғанымызда түркі тілдерінде тарихи корпус қарқындап дамымай жатқанын аңғардық.

Алдағы уақытта жасалатын қазақ тілінің тарихи корпусына бірнеше ұсыныстарымыз бар: 1) армян тілінің тарихи корпусынан байқағанымыз «Галерея» деген айдарында корпус бойынша жұмыс істеген адамдар мен жұмыс үдерісінің суреттерін түсіріп салған, мәтіндік ақпарат берілген; 2) Қазақстандағы кітапханалардың сирек қолжаба қорымен келісімге отыру; 3) Оқырмандарға тарихи корпустағы мәтіндерді, мысалдарды тегін көшіріп алуға мүмкіндік беру. Халықтың мұрасын шынымен де халыққа қолжетімді етуге атсалысу керек; 4) түрлі орфографиялық белгіленімдерді көрсету; 5) (<https://qazqorpus.kz/about/5/>) қазқорпустың сайтында көрсетілген жарияланымдардың pdf форматын енгізу қажет, бұл зерттеулерді жеңілдетеді; 6) «Басқа корпустар» деген айдарға туыс және туыс емес тілдердің корпустарына сілтеме бергенді жөн деп санаймыз. 7) Тарихи корпус жасау барысында «корпустық менеджер» деген лауазымды тағайындау қажет. 8) «Зерттеушілік бағдарлама» деген айдарда корпусты зерттеуге арналған мемлекеттік бағдарламалар мен жеңіп алған гранттар, үлес қосқан сайттар туралы ақпарат беру қажет деп санаймыз.

Бір сөзбен түйіндесек, қазақ тілінің тарихи корпусын жасау арқылы – тарихи сөздік, тарихи ономастика, тарихи терминология, тарихи орфография,

тарихи грамматика, тарихи сөзжасам, тарихи фразеология, тарихи тіркесімділік, тарихи теология, ескі діни мәтіндер мен тарихи жиілік сөздікке байланысты тарихи корпусты сала-салаларға бөліп, зерттеулерге жаңаша серпін беріп, ғылыми тұжырымдардың жүйелігіне, дәлдігіне, дәйектілігіне, зерттеулердің артуына негіз болады деп ой қорытамыз. Бұл мемлекеттік тілдің ғылыми базасының күшейуіне ықпал етеді. Бұл зор ауқымды міндетті орындау үшін қажырлы еңбек пен ұйымдасқан жүйе, мемлекеттің қолдауы қажет деп санаймыз. Тарихи корпусты жасау кез келген тілдің жаһандануға жұтылып кетпеуінің бір тетігі. Тарихи корпус кітапхана мен оқырманның арасындағы замануи цифрлық алтын көпір. Қазақ тілінің тарихи ұлттық корпусын жасау – мемлекеттік маңызды міндеттердің бірі.

4.3 Тарихи ішкорпус мәтіндерінің лексика-грамматикалық белгіленімі

Тарихи ішкорпус әзірлеуде белгілі бір графикамен жазылған қолжазбаның тек метасипаттамасы ғана берілмейді, сонымен қатар мәтіндегі әр лексикалық бірлікке лингвистикалық ақпараты бар белгіленімдер қойылады. Белгіленім – корпустың басты сипаттамасы. Сондықтан белгіленімнің түрлері неғұрлым көп болса, соғұрлым корпустың құндылығы артады.

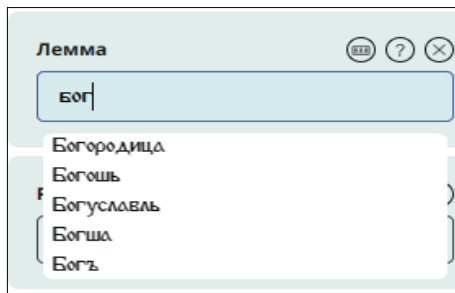
Жалпы белгіленімдер атқаратын функциясына қарай сыртқы және ішкі деп бөлуге болады. Сыртқы белгіленімге мәтін туралы ақпарат, яғни авторы, тақырыбы, жылы, жарны, стилі, көлемі, т.б. Оларды библиографиялық, типологиялық, тақырыптық, әлеуметтік, формальдық (мәтін, тарау, бөлім, абзац, сөйлем т.б.) және техникалық (орындаушылар, электрондық нұсқа алынған дереккөз, өңделген күні, кодталған кезі т.б.) деп те бөледі. *Ішкі белгіленімге* мәтіндегі әр сөзге берілетін лингвистикалық (фонетикалық, лексикалық, сөзжасамдық, синтаксистік, лингвомәдени, т.б.) ақпарат жатады [119]. Бұндай лингвистикалық белгіленімдері бар тарихи ішкорпустар, әсіресе, тілтанушы-зерттеушілер, мектеп мұғалімдері мен оқушылар үшін өте пайдалы ақпарат беретен дереккөз болатыны сөзсіз.

Тарихи ішкорпус жадына енгізілетін мәтіндердегі сөздерге лингвистикалық белгіленімдерді жүйелі құрылыммен қоюдың теориялық және практикалық әдіс-тәсілдерін меңгергенде ғана мүмкін болады. Осыған байланысты жалпы тарихи ішкорпус әзірлеу бойынша әлемдік қолданбалы лингвистика саласында құрылымы мен мазмұны тұрғысынан озық келе жатқан орыс тілінің тарихи ішкорпусын атауға болады. Аталмыш ішкорпус іштей көне орыс тілі, қабық хаттар, ескі орыс тілі және шіркеулік славян ішкорпустары болып бөлінеді. Барлық осы аталған ішкорпусының мәтіндеріндегі әрбір сөзге автоматтандырылған сөздің бастапқы тұлғасы (леммасы) мен грамматикалық белгіленімдері қойылған. Сондықтан зерттеуде орыс тілінің тарихи ішкорпус әзірлеушілердің, атап айтқанда, Т.С. Гаврилова, Т.А. Шалганова, О.Н. Ляшевская, С.О. Савчук, Д.В. Сичинова, сондай-ақ, қалмақ ғалымы Е.В. Бембеевтің, отандық ғалымдар А.Жұбанов, А.Жанабекованың, араб графикасынан кирил графикасына транскрипциялаған Р.Сыздық, Э.Фазылов, Э.Наджиб, Б.Сағындықов,

Б.Әбілқасымов, С.Иванов, Г.Сердюченконың еңбектері басшылыққа алынды.

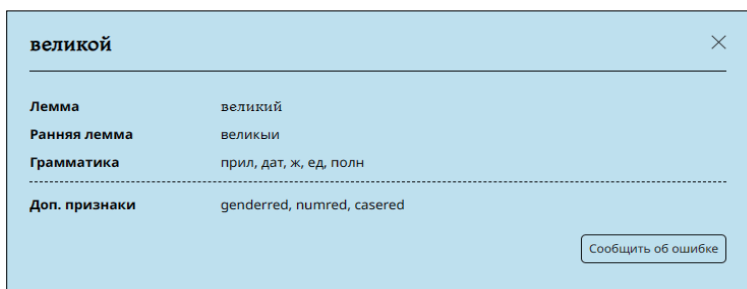
Ескі тілде жазылған мәтіндерге лингвистикалық белгіленімдерді әзірлеу – аса үлкен жауапкершілікті қажет етеді. Орыс тілінің тарихи корпусын әзірлеушілері ескі орыс тілінің мәтіндеріне автоматтандырылған лингвистикалық белгіленімдерді қоюда бірқатар мәселелерді атайды. “Біріншіден, мәтіндердің көбі XI-XIV ғғ. тілдің ерекшеліктерін көрсететіндіктен, мәтіндерде шіркеу славян тілінің лексикалық және грамматикалық элементтері мен әртүрлі аймақтарда жазылған мәтіндердің диалектілік ерекшеліктері де кездеседі. Бұндай ерекшеліктерді көрсету үшін компьютерлік бағдарлама «гибридті» парадигмаға негізделуі керек, яғни, ескі орыс және қазіргі орыс тілдерін ғана емес, шіркеу славян лексемаларын ескеру керек. Екіншіден, ескі орыс тіліндегі мәтіндердің емлесі тұрақсыз. Сондықтан корпус мәтіндерінде сөздің бірнеше орфографиялық варианттары болады. Мұндай варианттар бастапқы мәтіннің құрылуына немесе мәтін жасалған аймақтың диалектілік ерекшеліктеріне байланысты болуы мүмкін. Осыған орай лингвистикалық белгіленімдерде орфографиялық варианттар неғұрлым көп ескерілсе, мәтін соғұрлым толық және дәл талданады. Үшіншіден, техникалық қиындық – ескі орыс мәтіндерін автоматты түрде белгілеу үшін электрондық ресурстардың болмауы. Қазіргі орыс тіліндегі мәтіндерді өңдеу үшін жасалған анализаторлар ескі орыс мәтіндеріне жоғарыда аталған мәселелерге байланысты лексикалық бірліктердің гетерогенділігі мен орфографиялық варианттарды көрсету үшін қолайсыз болып отыр» [120].

Сонымен, көне орыс тілі, қабық хаттар, ескі орыс тілі, шіркеулік славян ішкорпустарының лексика-грамматикалық іздеу жүйесінің құрылымы өзара ұқсас болғанмен, айырмашылықтары да бар. Мәселен, көне орыс тілі ішкорпусы лексема, түсіндіру, грамматикалық, қосымша белгіленімдерден тұрады. Қабық хаттар ішкорпусы көне мәтіннің түпнұсқа және аудармасы, лемма, бастапқы лемма, сөзқолданыс, грамматикалық және қосымша белгіленім, түсіндіру, негізгі (орыс тіліндегі аудармасымен) семантикадан тұрады. Ескі орыс тілі ішкорпусы лемма, көне лемма, сөзқолданыс, грамматикалық және қосымша белгіленімдерден тұрады. Шіркеулік славян ішкорпусы лемма, сөзқолданыс, грамматикалық және қосымша белгіленімдерден тұрады. Осы аталған ішкорпустардың барлығына ортақ іздеу жүйесі грамматикалық және қосымша белгіленімдердің болуы. Грамматикалық белгіленімде лексеманың немесе сөзформасының қажетті морфологиялық сипаттамалары көрсетіледі. «Тандау» сілтемесі арқылы ашылатын қызмет терезесінде грамматикалық категорияларға бөлінген морфологиялық сипаттамалардың тізімі көрсетілген: сөз табы, септіктер, тегі, егіс, сан есім және т.б. «Қосымша белгіленімде» көрсетілген белгілердің көмегімен сөздерді белгілі бір позицияда іздеуге болады. Мәселен, тыныс белгісінің алдында және одан кейін, сөйлемнің басында немесе соңында. «Түсіндіру» іздеу жүйесінде омонимдерді ажыратуға мүмкіндік беретін көмескі мағынадағы сөздердің қысқаша түсіндірмелері көрсетілген. Сондай-ақ, сөздің бастапқы формасы (лемма) іздеу жүйесінде таңдалған сөздің лексемасы немесе сөз формасы көрсетіледі.



Сурет 16. «Бог» сөзінің леммасы мен сөзформасы

Лексикалық-грамматикалық іздеу жүйесі сөздің бастапқы формасынан (лемма) бөлек грамматикалық (сөз табы, септік, жеке, көпше түрі) сипаттамаларын анықтаудан тұрады. Пайдаланушы мәтіндегі кез келген сөзге меңзермен нұсқағанда ұяшықтан сол сөз туралы лингвистикалық ақпараттар ала алады. Айталық, «великий» сөзін меңзермен басқанда, мынадай грамматикалық (прил, дат, ж, ед. полн) анықтама беріледі.



Сурет 17. «Великий» сөзінің грамматикалық анықтамасы

Қалмақ тілтанушылары ескі қалмақ жазба ескерткіштері мәтіндеріндегі «көне» графемаларды өңдеу барысында UNICODE стандарттарына сәйкес кейбір символдарды кодтау мүмкін еместігін айтады. Тағы бір мәселе – сөздерді жазудың фонетикалық принципі, яғни сөздер қалай айтылса, жазба ескерткіштерде солай жазылған. Бір жағынан, бұл сол кезеңдегі қоғамның сөйлеу ерекшелігін көрсетеді. Бұл, әсіресе, лингвистер үшін маңызды. Ал бағдарламашылар үшін белгілі бір сөздің бірнеше варианттылығы лексикалық бірліктің танымдылығын көрсетуде қиындық туғызады [121].

Кейбір еуропа тілдерінің жазуы латын графикасында жазылғандықтан, тарихи корпустарының лингвистикалық белгіленімдері қиындықсыз әзірленген болса керек. Мәселен, португал тілінің тарихи корпусы 88 мәтінді лингвистикалық аңдатапасының жүйесі екі кезеңнен тұрады: 58 мәтіндегі сөздерге сөз табының белгісөзі қойылған, 27 мәтінге синтаксистік белгісөзі қойылған [122]. Көне грек корпусы цифрландырылған 820 көне грек мәтіндер базасынан тұрады. Корпус мәтіндеріндегі әрбір сөзі морфологиялық ақпараттармен қамтылған [123]. Ескі венгер корпусы XII-XVII ғғ. мәтіндерін қамтиды. Мамандар

мәтіндерді өңдеуде екі принципті басшылыққа алған: 1) қазіргі венгер тілінде қолданылмайтын барлық сөздерді морфологиялық ерекшеліктерімен сақтаған; 2) фонологиялық және орфографиялық ерекшеліктері бар барлық сөздерді алып тастаған. Яғни, әртүрлі нұсқада жазылатын сөздер қазіргі венгер тіліне жақындатылып, біріздендірілген. Бір мағынаны білдіретін екі түрлі жазылатын сөздер кездескен жағдайда «Венгер тілінің тарихи-этимологиялық сөздігіне» сүйенген. Мәселен, «бірақ» деген мағынаны білдіретін кедиг пен педиг шылаулары екі түрлі сөз ретінде жазылған [124].

Ал қазақ тілінің тарихи ішкорпусы мәтіндеріндегі жекелеген лексемаларға лингвистикалық белгіленімдер мен іздеу жүйесі қалай әзірленеді?

Орта ғасыр жазба ескерткіштері араб графикасымен жазылғанын ескерсек, арабграфикалы мәтіндерге лингвистикалық белгіленімдердің қойылуы өте қиын. Себебі араб графикасымен жазылған мәтіндердің қолжазбасы әртүрлі каллиграфиямен берілетіндіктен, іздеу жүйесінде таныла бермеуі мүмкін. Өйткені мәтіннің танылуында мынадай қиындықтар кездеседі: кейбір диакритикалық белгілердің (кесра, сукун, фатха), графикалық символдардың анық көрінбеуі немесе түсіп қалуы; кейбір сөздердің бірігіп жазылуы. Сондықтан жазба ескерткіштер мәтіндеріндегі кез келген сөз туралы лингвистикалық ақпарат көрсететін конкордансты жасау қиындық тудыруы мүмкін. Корпусқа енген арабграфикалы жазба ескерткіштер мәтіндерін компьютерде термейінше, скан (көшірме) нұсқасына лексикалық немесе грамматикалық белгіленім қою мүмкін емес. Сондықтан алғашқы уақытта араб графикасынан кирил графикасына транскрипцияланған нұсқаларына лингвистикалық белгіленім қоюға болады. Алайда араб әліпбиіндегі дауысты дыбыстардың үш-ақ түрі болғандықтан, түркі тілдеріндегі барлық дауыстыларды үш таңбаға сыйдыру мүмкін бола бермейді. Арабграфикалы жазба ескерткіштер мәтіндерін транскрипциялауда белгілі бір лингвистикалық нормаға бағыну қажеттігін В.В. Радлов, В.Л. Васильев және К.Г. Залеман, И.Ю. Крачковскийдің ұсынған кирил түркологиялық әліпбиінен бастау алады [125]. Десек те, татар, өзбек т.б. түркітанушы ғалымдар ескерткіштер мәтіндерін кирил таңбасымен бергенде өз тілдеріне жақындатып транскрипциялайды. Мәселен, өзбек ғалымы Э.Фазылов Хорезм мәтіндеріне отандық түркологтардың белгілер жүйесін пайдаланады, алайда өз тарапынан аздап толықтырулар жасап транскрипциялайды [126]. С.Н. Иванов Әбілғазының «Түркі шежіресін» транскрипциялағанда академиялық (радловская) транскрипцияны пайдаланады [127]. Аталмыш жазба ескерткішті Б.Әбілқасымов араб тіліндегі дауысты дыбыстарды қазіргі қазақ тілінің дыбыстарына салып транскрипциялайды [128]. Р.Сыздық та Йасауи қолжазбасына транскрипциялауда қазақ тілінің фонетикалық жүйесі мен дыбыстық заңдарына (мысалы, ы-і, а-ә, ұ-ү қатарларын айырып, дауыстылардың жуан-жіңішкелігін, яғни дауыстылар үндестігін сақтап көрсету) сәйкестендіре транскрипциялайды [129].

Ғалымдардың транскрипциялаған әр ғасырдың мәтіндерінен үзінді келтірейік. XI ғасырдағы Ахмет Иассауидің «Диуани хикметі» Самарқанд (Залеман) қолжазбасының транскрипцияланған нұсқасынан үзінді:

Күлбары хали / йігірмә бешкә йетмәй
Маһу сали/ насихатлар қылар пир-у жууанны/
Өзі фәһм етмәгән йахшы йаманны/
Қуруб дамйн кәззаблар кәни би пир/ тілі
Мәкр-у хилә қылғаны тасхир/ аны тәзуир
Дүр шайтандын а'лә/ копр йүзі қара

XII ғасырдағы Ахмет Йүгнекидің «Хабатул хақайық» шығармасынан үзінді:

Илаһи өкүш хамд айурмен сеңа
Сениң рахматиндин умармен өң-а
Сана му айуғай бу тилим
Унарча айайын йары бер меңа
Сениң барлығыңға тануғлуқ берүр
Жумат жануар учған йүгүрген нең-а
Сениң бирлығыңға далил арқаған
Булур биър нең ичра далиллар мың-а.

XIV ғасырдағы Сейф Сарайдің «Гүлістан бит-түрки» шығармасынан үзінді:

Алтун қанатын ачты есә субх шунқары
Көк көлгә батты жумлә кәвакиб кәбутәри.
Әсфәндәри Рум чықып тиксә әләм
Болуп көруп хәзимәт аны Шам ләшкәри.
Төкти қырмызы қан, оқытып чирх тасына
Туннуң йурәк тамырын ачып, субх шиштәри
Қушлар көруп хавада аны башлады раван

XVII ғасырдағы Әбілғазы баһадүр ханның «Түркі шежіресі» шығармасынан үзінді:

Артұқ бөке бин Толы хан нәсілінден ерді хан көтерділер Ғали ойратке Бағдад хәкімі ерді Мосы бин Ғали бин Байду бин Тарағай бин Һалақу ханны хан қылдылар Арба хан бірлән ұрұшыб ғалыб келді тақы Арба ханны өлтүрді мемлекетіні тахт тәсарафға кіркүзді ол уақытта Шейх Хасан Жалайрке аңа Шейх Хасан Бузруг дерлер Румның хәкімі ерді ол Мосы болғанын ешітіб Махмүд бин йол құтлұқ бин Тимур бин Анбаржу бин Менку бин Һалақу ханны хан қылыб Румдын уа Кәржиден ләшкері ғазім жамиг қылыб Иран тарафыға мәтаужаһ болды.

Берілген үзінділерден көріп отырғанымыздай, авторлар бірқатар сөздерді әртүрлі транскрипциялаған. Мәселен, XI ғасырдағы Ахмет Иассауидің «Ди-уани хикметінде» қыпшақ тіліне тән ы-і, а-ә, ұ-ү дыбыстарының үндестігі барлық уақытта сақталмаған. XII ғасырдағы Ахмет Йүгнекидің «Хабатул хақайық» мәтіні ы-і, а-ә, ұ-ү дыбыстар үндестігімен транскрипцияланбаған.

XIV ғасырдағы Сейф Сараидің «Гүлістан бит-түрки» мәтiнiнде сөздердiң жiнiшкелiк үндестiгiн сақтау үшiн ä ö сияқты умлаут белгiсi қолданылған. «Жамиғ-ат тауарихта» мәтiнiнде ь графемасы брефиспен ў, ʹ графемасы макронмен ā таңбасымен транскрипцияланған. Мысалы, اولايت – ўилайат, т.б. Араб сөздерiндегi ڭ (айын) дыбысы кейбiр мәтiндерде апострофпен ‘, кейбiр мәтiндерде ғ таңбасымен транскрипцияланған. Мысалы, та‘ ала – тағала, ‘ақл – ғақыл, т.б.

Қарахандықтар және Алтын Орда дәуiрi жазба ескерткiштерi мәтiндерiнде өзбек әдеби тiлiнiң дыбыстық ерекшелiгiне сәйкес транскрипциялау басымдық алған. XVII ғасырдағы Әбiлғазы баһадүр ханның «Түрki шежiресiн» транскрипциялауда өзге шығармаларға қарағанда дыбыстардың жуан-жiнiшкелiгi ескерiлген. Осыған байланысты жазба ескерткiштер мәтiндерiнде бiр мағынадағы сөздiң әртүрлi транскрипцияланған варианттары пайда болған. Оларды графикалық, фонетикалық, грамматикалық варианттар деп бөлуге болады. Бұндай ерекшелiк орыс тiлiнiң корпусында да кездеседi.

«Орфографиялық нормасы мен оның варианттары: естли {если*}, зделать {сделать*}, довольно {довольно*}» [130].

Кесте 6

Жазба ескерткiштер мәтiндерiндегi вариант сөздер

Орфографиялық вариант		Фонетикалық вариант		Грамматикалық вариант	
норма	вариант	норма	вариант	норма	вариант
бiрлән	бiрлә/бiлән	Чыңғыз	Шыңғыз	екiнiң	Екi нiң
йол	иол	алтұн/алтун	алтын	бiрiсi	Бiрi сi
барұб	баруп/барұп	деб	теб	һәр түрлү	һәр түрлү
негоһ	неках/некох	дағы	тағы	сусулықдан	сусулықдан
бесиор	бесиар	мұндағ	мұндақ	хабарлы	хабарлы
маңа/баңа	меңа	тiрiк	тiрi	ачлықдан	ачлықдан
саңа	сеңа	құтлығ	құтты	қамашлар	қамашлар
аңа	анга	уа	ва	һауаға	һауаға

Орфографиялық варианттар араб тiлiндегi үш созылыңқы және үш қысқа (фатха, кесра, дамма) дауысты дыбыстарын кирил графикасымен транскрипциялауда дауыстыларды дәлме-дәл бере алмайды. Сол сияқты ڭ (ж), ڪ (к), ڤ (б) дауыссыздары жазба ескерткiштер мәтiндерiнде екi-үш дыбыстың функциясын атқарған. Мәселен, ڭ (ж) әрпi бiрде ж, бiрде ч, бiрде ш дыбыстарының функциясын, ڪ (к) әрпi бiрде к, бiрде г және ғ дыбыстарының функциясын, ڤ (б) әрпi сөздiң соңғы позициясында бiрде б, бiрде п графемаларымен таңбаланған. Сондай-ақ орфографиялық варианттар түрiн көбейтетiн кiрме араб, парсы сөздерiн әртүрлi нұсқада транскрипциялауда көрiнiс табады: «көп» сөзiнiң синонимi ретiнде парсы сөзi бiрде бесиор, бiрде бесиар, бiрде бисйар

деп транскрипциялана берген. Орта ғасыр жазба ескерткіштерінің тілі аралас тілде жазылғандықтан, қыпшақтық, қарлұқтық, оғыздық белгілері мен архаикалық элементтеріне қарай фонетикалық варианттар да пайда болған. Кейбір мәтіндерінде қосымшалар түбір сөзден бөлек жазылған. Оның себебі араб графикасында өзінен жалғастырып жазбайтын жеті таңба бар. Сондықтан болса керек, мәтінде қалай жазылса, солай транскрипцияланған. Осындай тұрақты орфографиялық норма болмағандықтан, транскрипциялауда да ала-құлалық тудырған. Сондықтан ортағасыр транскрипцияланған мәтіндердің танымдылығын арттыруда лингвистикалық белгіленімдерінің графикалық, орфографиялық, фонетикалық варианттарды ескеруіміз қажет. Яғни, олардың түрлі варианттарын корпустың лингвистикалық белгіленім жүйесіне енгізуіміз қажет. Осылайша корпуста лексика-грамматикалық іздеу жүйесінде варианттардың барлық түрі беріледі. Сондай-ақ, әртүрлі нұсқада транскрипцияланған сөздердің қай варианты басым екендігін анықтауды қажет етеді. Лингвостатистикалық талдау нәтижесінде ең көп кездесетін орфографиялық варианттар – а мен ә, ы мен і, у мен ұ дыбыстарды таңбалайтын графемалар аралығында. Осындай транскрипциялаудағы ала-құлалық, әсіресе, қарахандықтар мен Алтын Орда дәуіріндегі жазба ескерткіштер мәтіндерінде көбірек кездеседі.

Сонымен, араб жазулы жазба ескерткіштеріне лингвистикалық белгіленім жүйесін арабграфикалық мәтіндеріне емес, транскрипцияланған мәтіндеріне қойылады. Лингвистикалық белгіленім лексикалық және грамматикалық белгіленімдерін қамтиды. Корпусқа енетін жазба ескерткіштер мәтіндеріне лексикалық және грамматикалық белгіленім мынадай ретте көрініс табады:

Сөзтізбе. Транскрипцияланған мәтіндегі әрбір лексикалық бірлік сөзтізбеге алынады. Онда сөздің негізгі нұсқасы беріледі.

Вариант. Мәтіндегі бір мағынадағы әрбір сөздің түрлі нұсқадағы көрінісі. Сөзтізбедегі орфографиялық, фонетикалық варианттары беріледі. Сөздердің варианттары неғұрлым көп берілсе, соғұрлым іздеу жүйесінде танылуына мүмкін болады.

Лемма. Сөздің бастапқы тұлғасы, түбірі беріледі.

Лексикалық қабаты. Лексикалық бірліктің этимоны беріледі.

Мағынасы. Лексикалық бірліктің мәнмәтіндегі мағынасы беріледі.

Сөз табы. Лексикалық бірліктің сөз табына қатыстылығы анықталады.

Тұлғасы. Негізгі және туынды тұлғасы анықталады.

Қосымша. Көптік, тәуелдік, септік, жіктік жалғаулары мен сөзтудырушы жұрнақтары анықталады.

Арабграфикалы мәтіндерге арналған лингвистикалық белгіленімдер бағдарламасының жұмыс істеу механизмдерін қалай әзірлейміз? Тегтер дегеніміз – белгілі бір мәтіннің лексикалық-грамматикалық аннотациясын автоматты түрде жүргізуге қабілетті бағдарлама. Лексикалық-грамматикалық аннотациясын автоматты түрде енгізуде мынадай бағдарламалар қолданылады:

– Google кестесі;

– Python бағдарламалау тілі;

– Django Web сервер бағдарламасы;

– MySQL мәліметтер қорын басқару жүйесі.

Google кесте онлайн бағдарламасы – мәтіндерді жинақтауға және оларды теңестіруге ыңғайлы құрал. Бұндай жиналған мәтіндер Google кесте файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктеуге де жеңіл. Сонымен қатар Notepad++ мәтіндік процессор бағдарламасы қолданылады. Бұл бағдарлама тарихи мәтіндерді сұрыптауға және статистикасын алуға көмектеседі. Осылайша тарихи ішкорпусты интернетке жариялау үшін Django вебсервері қоланылады. Бұл сервер қолданушының сұранысын қамтамасыз етеді. MySQL мәліметтер қорын басқару жүйесімен байланыстыратын ортақ Processing.py бағдарламасы жасақталды. Бұл бағдарлама Python тілінде жазылған және тарихи ішкорпусының іздеу функцияларын атқарады. Осылайша ішкорпустың әр сөзіне лингвистикалық паспорт әзірленді. Оған арнайы паспорт форматы мен бағдарламасы жасақталды. Ішкорпустың лингвистикалық паспортының сұлбасы 18-суретте көрсетілген.

Сөзтізбе	Сөз	һәр	тұрұрлар
	.Лемма	һәр	тұр
	Вариант	хәр	турурлар
Лексикалық белгіленім	.Лексикалық қабаты	парсы	төл
	Мағынасы	әр	тұр
Грамматикалық белгіленім	Сөз табы	есімдік	етістік
	Тұлғасы	түбір	негізгі
	Қосымша	-	-лар: көптік жалғау

Сурет 18. Тарихи ішкорпустың лингвистикалық паспорты

Паспортта әр тарихи сөздің сөз, лемма, вариант, лексикалық қабаты, мағынасы, сөз табы, тұлғасы, қосымша сипаттары (разметка) көрсетіледі. Бұл сипаттардың бәрі алдын ала анализатордан өткізіліп, мәліметтер қорын басқару жүйесіне де жылдам іздеуге ыңғайлы етіп сақталады. Жалпы ішкорпустың бағдарламалық құрылымы 19-суретте көрсетілген.



Сурет 19. Тарихи ішкорпустың бағдарламалық құрылымы

Тарихи ішкорпусының бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологияларын пайдалана отырып, әзірленді.

Мұндағы Model – тарихи ішкорпустың мәліметтер қорын жинақтайтын жүйе. View – пайдаланушыға жұмыс істеуге арналған интерфейс сұлбасы. Controller іздеу функцияларын орындауға және View мен Model-ді өзара біріктіруге қызмет етеді.

Қорыта келгенде, тарихи ішкорпусқа араб графикасынан кирил графикасына транскрипцияланған мәтіндерге лингвистикалық белгіленімдерді енгізу үшін төмендегідей зерттеулер жасалды:

- корпусқа еніп отырған XI-XVII ғғ. жазба ескерткіштері мәтіндерінің араб графикасынан кирил графикасына транскрипцияланған мәтіндерді саралағанымызда, бір кезеңге жататын мәтіндердің өздері әртүрлі транскрипцияланған. Олардың ішінде Қарахандықтар және Алтын Орда дәуірі жазба ескерткіштері мәтіндері өзбек әдеби тілінің дыбыстық ерекшеліктеріне сәйкес транскрипцияланған;

- транскрипциялауда әртүрлі орфограмманы тудырған дауысты, дауыссыз дыбыстардың негізінде орфографиялық, фонетикалық, грамматикалық варианттар анықталды;

- лингвостатистикалық талдау нәтижесінде ең көп кездесетін орфографиялық варианттар – а мен ә, ы мен і, у мен ұ дыбыстарды таңбалайтын графемалар;

- транскрипцияланған мәтіндер бойынша лексика-грамматикалық белгіленімдер жүйесі әзірленді;

- транскрипцияланған мәтіндерге арналған лингвистикалық белгіленімдер бағдарламасының жұмыс істеу механизмдері сипатталды.

Сонымен, тарихи ішкорпусқа енетін мәтіндерге лексикалық, семантикалық белгіленімдердің қойылуы, талданған мәтіндердің көпшілікке қолжетімді болуы – диахрондық лингвистиканы сапалы зерттеуге мүмкіндік береді.

5 ПАРАЛЛЕЛЬ ІШКОРПУС

5.1 Аударматану ғылымының басты ұғым-түсініктері

Қазақ мәдениетіндегі аударма ісінің алғашқы үлгілері Абай Құнанбайұлының, Ахмет Байтұрсынұлының шығармашылығында кеңінен орын алғандығы мәлім [131]. Ал қазақ тілтанымында аударманы лингвистика ғылымының аспектісінде зерделеу ісі қазақ тілін теориялық тұрғыдан танып-білудің негізін салған белгілі ғалым Құдайберген Жұбановтың еңбектерінен басталады. Ол өзге тілден қазақ тілін тәржімалау ісіндегі синтаксистік құрылымдардың қандай ерекшеліктері болатындығын, стильдік айырмашылықтарды талдады [132, 531].

Өткен ғасырдың орта тұсында атақты тұлға Мұхтар Әуезовтің пікірлері осы саладағы басты мәселелерді айқындап берді. Жазушы тәржіма ісі саласындағы нақты міндеттерді шешу үшін аудармадағы ұлттық дәстүр, ұлттық форма қалай берілуі керек, синонимдер, идиома, мақал-мәтел, афоризмдер қалай аударылған, аудармашы өз тарапынан қандай-қандай жаңа сөз туындыларын қосты, жаңадан қандай сөз байланыстарын жасады, төл сөздердің мағыналарын қаншалықты тереңдетіп, жаңартты, – міне, осы аталғандарды зерттеу қажет деп санап, аудармаға тілтанымдық тұрғыдан келудің маңыздылығына баса назар аудартты [133, 7].

Содан бергі кезеңдерде, әсіресе соңғы ширек ғасыр көлемінде аударманың нысанасы да, ауқымы да ұлғайды. Алдыңғы тұстарда орыс тілінен қазақ тіліне көркем әдебиетті тәржімалау ісі белсенді жүргізілсе, кейінгі жылдары мәдениетаралық коммуникация өрісінің кеңеюіне байланысты дүниежүзілік ақпараттар алмасуды барынша сапалы деңгейге көтеру мүддесінде еуропалық тілдерден, шығыс тілдерінен қазақ тіліне тікелей тәржімалау үдерісі қанат жая бастады. Ақпараттық-анықтамалық база құру үшін публицистикалық, қоғамдық-саяси мазмұндағы материалдар, нормативті құжаттар аударма арқылы жасалуда. Қазақстан Республикасының алыс және жақын шетелдермен қазіргі әлеуметтік-мәдени байланыстарының жаңаша, қарқынды арнада дамып отырғандығына қарай маман-аудармашылар даярлау ісіне де баса көңіл бөлінуде. Ал болашақ мамандар, сондай-ақ тәржіма ісінде өз шығармашылығын байқап көріп, дамытып жүргендер үшін М.Әуезов айтқандай, «аударманың бірбеткей көлеңке жақтарын... байқастыру» жеткіліксіз; аудармашының кәсіби біліктілігінің жоғары деңгейде қалыптасуында түпнұсқа тіл мен аударылатын тілдің типологиялық-құрылымдық ерекшеліктерін, мазмұны мен құрылымы дайын түпнұсқа мәтінді тәржімалау үдерісінде қазақ тілінің лексикалық-грамматикалық әлеуетінің қандайлық дәрежеде қозғалысқа түсетіндігін жетік меңгеру қажеттігі туады. Сонымен қатар аудармашы аударудың негізгі ұғымдары мен түсініктерін, аударудың жолдары мен тәсілдерін де білуде де біліктілік танытуы жөн.

Әлемдік лингвистикадағы лингвистикалық аударматану ғылымының басты мәселелерінің бірі – осы салаға қатысты терминдер мен терминдік сөз тіркестері. Орыс тіл білімінде аудармаға қатысты екі мыңнан астам аталымдар

бар екендігі көрсетіледі [134, 2003], ал қазақ тіл білімінде олардың көпшілігі құрылымдық тұрғысы жағынан орыс тілінен сөзбе-сөз аудару арқылы орныққан. Дегенмен, шынтуайтқа келгенде, білім беру ісінде, кейбір мәселелерді принципті түрде талқылауда аударма ісінің кейбір терминдері ұғымдық мәні, мазмұндық межесі жағынан нақты емес. Біздің мақсатымыз – аударматану ғылымының бірқатар өзек аталымдарына – терминдеріне орыс тілінен қазақ тіліне аударылған мәтіндердің заңдылықтарына сүйене отырып, мүмкіндігінше анығырақ түсініктеме жасау.

Жазбаша аударма – кәсіби (және әуесқойлық) аударманың кең тараған түрі; қағазға түсірілген жазбаша мәтінді екінші тілге жазбаша түрде аудару және бұдан алынған нәтиже. Түпнұсқа мәтін алдын ала танысу, көру, ойлау, пайымдау, талдау, зерттеу, сөздіктерді, басқа да лингвистикалық және лингвоелтанымдық, когнитивтік мағлұматтарды беретін анықтамалықтарға иек арту арқылы екінші тілге жазбаша түрде түсіріледі. Бұл үдерістің өтуі барысында аудармашының мүмкіндігі шексіз; ол екінші тілде жасалған мәтінді өңдеп, реттей алады. Жазбаша аударма мәтін тәржімашы шығармашылығының нәтижесі болып саналады. Жазбаша мәтінді ауызша аударып беруді жазбаша-ауызша аударма деп атауға болады. Қазақ мәдениетінде классикалық көркем әдебиеттің нұсқалары, ғылыми әдебиет және ресми, нормативті құжат үлгілері жазбаша түрде аударылады.

Ауызша аударма – бұл тәржіманың жазбаша түріне қарағанда күрделірек құбылыс; ақпараттың мазмұн межесін (план содержания) және құрылым межесін (план выражения) жазбай, ауызша жеткізу түрі. Аударматану бойынша ғылыми еңбектерде ауызша аударманың екі түрі аталады:

- ілеспе (синхронный) аударма;
- жалғас (последовательный) аударма.

Ілеспе аударма – бастапқы мәтінді тындап қабылдай отырып, ілесе, қосарласа, бір мезгілде аудару. Бұл үдеріс соңғы жылдары арнаулы жабдықтар арқылы жетілдірілген. Аударудың бұл түрі түпнұсқа мәтіннің айтылуымен қатар жүріп отырады. Арнаулы кабинеттерде ілеспе аударма түпнұсқаның мәтінін сөйлеушіге (мәтіннің авторына) ілесе отырып оқу арқылы, кей жағдайда сөйлеушінің ауызша баяндалған сөзін алдын ала дайындықсыз тікелей аудару арқылы жүзеге асырылады. Ауызша аударманың қай түрі болсын, аудармашыдан жоғары кәсіби біліктілікті, тәжірибенің болуын, екі тілден де лингвистикалық тұрғыдан құзыретті болуын талап етеді. Аударманы ауызша жасау кезінде уақыт тапшы, түпнұсқа мәтін бір-ақ рет айтылады немесе оқылады, сөйлеушінің (мәселен, баяндама жасаушының) өзіне тән дауыс ырғағы, сөйлеу мәнері, екпіні болады; демек, алынған ақпаратты тез қорытып, тыңдаушыға екінші тілде кедергісіз түсінетіндей (бұл ең бастысы десек те болар), әрі сауатты жеткізу міндеті бар. Ауызша аудармадағы тағы бір күрделі тұс – аудармашының санасы бір тілден екінші тілге шапшаң ауысып тұрады, түпнұсқадағы бір сөйлемді аударып жатқан кезде түпнұсқаның екінші сөйлемі басталып жатады, сондықтан маман түпнұсқаның тыңдау арқылы қабылдануын бөлік-бөлшектерге бөле алатындай тәжірибе жинақтауы қажет. Ауызша аударма кезінде тұтас құрылымды сегменттер арқылы жеткізуге бола-

ды, бұл әсіресе әр басқа типологиялық ерекшеліктерге ие орыс тілі мен қазақ тілдеріне тән. Біздің байқауымызша, ілеспе аударма кезінде сөздің таңбалық құрылымы, абсолютті сәйкестіктер (сөздіктегі денотат баламалар) алынады, өйткені түпнұсқадағы тілдік бірліктің коннотатты мағыналарын, экспрессивті реңктерін игеруге уақыт жібермейді. Орыс тілінен қазақ тіліне ілеспе аударма жасау кезінде қазақ тілі сөздерінің мағыналық байлығын, тұрақты сөз тіркестерін еркін қолдана білсе, бұл аудармашының лингвомәдени құзыретінің жоғары деңгейде екенін айқындайды. Алайда ықшамдау, кейбір мәліметтерді түсіріп тастау – ілеспе аударма кезінде жиі кездесетін, бірақ барлық жағдайда да дұрыс деп бағаланбайтын тәсілдер.

Жалғас аударма – түпнұсқада сөйлеушінің ақпаратын тыңдап алғаннан кейін жүргізілетін амал. Жалғас аударма, әдетте, мінберден сөйлеп тұрған шешеннің бір абзацты, бір фразаны немесе бір сөйлемді айтып болғанынан соң жүзеге асырылады. Бұл реттегі қиындық – түпнұсқада сөйлеуші өз ойын қай шамада тоқтатамын десе де өз еркінде. Тілдік құрылымы күрделі, ақпараты ауқымды бір абзацтан кейін пауза жасалған жағдайда жалғас аударманы жүргізуші айтылғандарды жадында сақтап, ақпаратты екінші тілде мүмкіндігінше толық қамтуға міндетті. Аударманың бұл түрінде ықшамдау, функционалдық қызметі бәсең ақпаратты түсіру, орнын ауыстыру тәсілдері игеріледі, өйткені айтылған ақпарат адам жадының оперативті мүмкіндігіне сыймауы мүмкін, әрі түпнұсқада сөйлеушінің келесі сөзі басталып та кетуі мүмкін.

Қазіргі кезде жалғас аударма ағылшын тілінен қазақ тіліне тікелей тәржімалау ісінде байқалады.

Ілеспе аударма және жалғас аударма – баяндамаларды, қоғам қайраткерлерінің сөздерін, мәжілістер мен мәслихаттарда мінберден айтылған, оқылған сөздерді тез арада ауызша жеткізу үшін жүзеге асырылатын қызмет түрі.

Ескерте кетерлік жайт – қазіргі әдеби тілде ілеспе аударма және қосарласа аударма деген атаулар қолданылып жүр; олар синхронный перевод, последовательный перевод деп аталатындардың орнына әрқилы жұмсалады. Ілеспе және қосарласа деген сөздердің семантикасы қатар, ілесе жүретін үдеріс дегенді білдіретіндіктен, ал орыс тіліндегі последовательный перевод атауы айқын түрде, түпнұсқа мәтіннен кейін жүзеге асырылатын, түпнұсқа мәтінге жалғастырыла орындалатын тәржіме қызметін нақты ұғындыратынына байланысты біз терминдер ұғымында екіұдайылық болмауын көздеп, соңғысына жалғас аударма деген аталымды ұсынып отырмыз.

Аударманың типтерін саралағанда, қазақ мәдениетіндегі аударма ісіне қатысты пікір-тұжырымдарда сөзбе-сөз аударма, дәл (дәлме-дәл), калька аударма, әріпқойлық деген ұғымдар жиі ұшырасады. Көрсетілгендер кейде нақты терминдік ұғымға ие аталымдар тәрізді болғанымен, ғылымның теориялық бағыттары тұрғысынан олардың анықтамаларын нақтылай түсуге тура келеді. Өйткені проблеманың осындай жақтарын барынша таратып айтқан орыс тілші ғалымдарының тұжырымдары да кейде біркелкі шықпай жатады. Оның да себептері жоқ емес: бір мәтіннің ішінде аударма бірліктерінің бірі сөзбе-сөз тәржімаланғанымен, сөйлемнің өзі тұтастай алғанда, еркін аударма түрінде көрінетін жайттар да болады.

Әріпқойлық – түпнұсқа тіл мен аударма тілдегі формалды немесе семантикалық құрылымды түп-тура сәйкестендірудің нәтижесінен туатын құбылыс. Бұл – дәлдік емес, бұл – түпнұсқадағы ақпараттың мағынасын бұзатын, кері түсіндіретін және аударылатын тілдің нормасына елеулі нұқсан келтіретін жайт; аудармашының қатесі немесе аудармашының лингвомәдени құзыретінің төмендігін көрсететін белгі. Әріпқойлық жеке сөзден гөрі сөз тіркесінің, фразалық тұтастықтың, сөйлемнің аудармасында айқын көрініс береді. Оның себебі мынада: аудармашы жеке сөзді алғанда оның сөздіктегі сәйкес баламасын игереді немесе контекстуалды, функционалды баламаларын қолданады, бұл аудармадағы заңдылық болып саналады. Ал сөз тіркесінен бастап былай қарай күрделене беретін құрылымдар үшін сөздіктегі сәйкестіктердің алынуы немес калька баламаның жасалуы әріпқойлыққа ұрындырады. Әріпқойлық – әсіресе көркем мәтіннің аудармасына нұқсан келтіреді.

Сөзбе-сөз аударманы қазақ ғылымында аударманың сапасын төмендететін, аударма теориясының әдістерін, техникасын меңгермеуден туатын құбылыс деп бағалаушылық бар. Бұл – негізінен көркем әдебиетті аудару практикасын зерттеуден туындыған қорытынды. Сөзбе-сөз аударма көркем мәтінді тәржімалаудағы әмбебап тәсіл емес. Бастапқы мәтіннің формалды немесе семантикалық компоненттерін (сөйлемнің орналасу тәртібін, қосымшалар мен түбір сөзді калька арқылы жасау) екінші тілде тура қайталау – сөзбе-сөз аударудың үлгісі. Көркем әдебиетті дәл осы жолмен тәржімалаған жағдайда ақпараттың мазмұнына, аударылатын тілдің әдеби нормаларына нұқсан келеді. Сонымен қатар сөзбе-сөз аударма – жанрлардың қайсыбір түрлерін аударуда қажетті амал. Айталық, нормативті құжаттар аудармасының төмен болуы түпнұсқаның – орыс тілінде жасалған мәтіннің – қазақша сөзбе-сөз тәржімаланғанынан деген пікірлер жиі айтылып келеді. Шындығында нормативті құжаттардың күрделі түрлері (жарғы, жарлық, қаулы, ереже және т.б.) сөзбе-сөз аударылады. Бұлай болуы – аудармашының лингвомәдени құзыретінің төмендігінен емес. Ресми құжаттарды сөзбе-сөз аударып беруге міндетті, мәтінді қазақы тілмен мазмұндап беруге, еркін аударуға, арасына сөз, сөз тіркестерін қосып нақтылап түсіндіруге, түсіріп тастау амалын қолдануға маманның құқығы жоқ. Өйткені бірнеше тілде жасалған заңнамалық құжаттардың барлығының да заң күші бірдей болуы керек. Аударылатын тілдің табиғи заңдылығына салып, «бейімдеп жіберу» ресми құжаттың заң күшін әлсіретеді, бір тілдегі мәтін бір мағынада, ал екінші тілдегі мәтін екінші мағынада ұғынылуы мүмкін. Сондықтан аудармашы нормативті құжаттарды екінші тілге аударғанда, сөзбе-сөз аударуға амалсыз, әрі заңға сүйене отырып, әділетті, сауатты түрде барады.

Сөзбе-сөз аударма калька ұғымымен ұштасады, кейде екеуі де бірдей ұғым-түсінік (яғни калька дегеніміз – сөзбе-сөз аудару) ретінде қарастырылады. Олардың мән-дес, ұғымдас түсініктер екендігі рас. Дегенмен көптеген зерттеулердегі көзқарастарды жан-жақты ескере отырып, төмендегілерді нақтылай түсу қажет болады:

Калька – сөзді немесе сөз тіркесін сөзбе-сөз аударудың нәтижесі; басқа тілдегі лексикалық-синтаксистік модельдерге қазақ тілінің модельдерін сәйкес келтіре отырып, жаңа сөз немесе сөз тіркесін жасау. Өзге тілдегі сөз тіркесінің,

сөздің бір бөлшегі үшін қазақ тіліндегі тура, сөздікте ұсынылатын, грамматикада көрсетілетін сәйкестік алынады. Мысалы, ағылшынша-орысша-қазақша сәйкестіктерді салыстырсақ, *House of Culture – Дом культуры – Мәдениет үйі* дегендер калька болып саналады. Калька арқылы жасалғандарға импорталмастыру (импортообмен), үштік одақ (союз трех), жасыл көше (зеленая улица), төртінші билік (четвертая власть), түзеткіш (выпрямитель), келтірілген ұзындық (приведенная длина) деген үлгілерді атауға болады. Алайда калька – формалды аударма; жиі қолданысқа түскен сөз және сөз тіркесі (терминдерден басқасы), калька үлгісінің «сериялы» түрде жұмсалуды тілдік стереотиптердің пайда болуына әкеледі (*салыстырыңыз: ақ алтын – қара алтын – жұмсақ алтын*). Сөйтіп, сөзбе-сөз аударма сөйлем, абзац деңгейінде, ал калька аударма сөз, сөз тіркесі деңгейінде болады.

Көптеген зерттеулерде дәлме-дәл аударма деген терминді кездестіреміз, бұның да калька, сөзбе-сөз аударма, әріпқойлық дегендермен синонимдес жұмсалатын реттері жоқ емес. Әр атауды нақтылай түсер болсақ, дәлме-дәл аударма – екінші тілде бар бірнеше сәйкестіктердің біреуінің түпнұсқа мәтіннің мазмұнына және ең бастысы – аударманы оқитын адамның ұлттық-танымдық қабылдауына, тілдік деңгейіне сәйкес келуі. Мәселен, «На улице идет снег» деген сөйлемнің бірнеше нұсқасын жасауға болады: Көшеде қар келе жатыр / Көшеде қар жауып жатыр / Далада қар жауып тұр.

Алғашқысы – калька компоненттерге құрылған сөзбе-сөз аударма; екіншісінде ауызекі сөйлеу тіліне тән ерекшеліктер бар. Соңғы вариант – түпнұсқаға толық сәйкес келетін дәлме-дәл балама аударма болып табылады; бұлай деуімізге себеп – қазақ ұлтының концептілік ұғымына, сыртқы дүниені қабылдау, пайымдауынан қалыптасқан тәжірибесіне негізделген тілдік бірліктер алынған (шынайы қазақы ұғым үшін тұрғынжайдан тысқарының бәрі «дала», сондықтан әрдайым «далада» деп айтылады), іс-әрекеттің дәл осы сәтте жүріп жатқандығын көрсететін тұлға (жауып тұр) жұмсалған. Лингвомәдени құзыретке ие аудармашы нұсқалардың үшіншісін ғана таңдайды, соңғы вариант түпнұсқадағы ақпараттың мазмұнына сәйкес, қазақ әдеби тілінің нормалары сақталған, қабылдаушының ұғымына сәйкес келетін дәлме-дәл балама балып бағаланады. Қорыта айтқанда, дәлме-дәл аударма жазбаша тәржімаға көбірек тән; аудармашы бірнеше варианттардың қатарынан дәлме-дәл баламаны таңдай алады.

Бұл термин орыс тіліндегі «адекватты» дегенге балама ретінде ұсынылып отыр, осы аталымды кейбір еңбектерде «сайма-сай аударма» деп беру де ұшырасады.

Еркін аударма дегеніміз, атаудың өзінен көрініп тұрғанындай, түпнұсқаның екінші тілде барынша еркін жасалуы. Бұл – субъективті құбылыс, яғни бір мәтінді әр аудармашы өзбетінше еркін түрде жасаса, олардың туындылары мазмұны сақталғанымен, құрылымы жағынан әр бөлек болады. Еркін аудармада құрылым сақталмайды, жалпы мазмұн қысқартылып немесе ұлғайтылып ұсынылады, бұнда трансформацияның барлық тәсілдері игеріледі. Бұл сипаттар қазақ аударматануындағы алғашқы еңбектердің бірінде жан-жақты қарастырылды (Әуезов, 1957). Көптеген ғалымдар еркін аударманы тәржіманың ең

жоғарғы деңгейі деп есептейді (Көркем аударма..., 1957). Бұл – көркем әдебиет жанрларын, әсіресе поэзияны тәржімалауда жиі игерілетін тәсіл.

Аударматану ғылымының өзекті терминдерінің тағы бірі – сәйкестік.

Аударма қызметінің лингвистикалық теориясының қалыптасуының алғашқы кезеңдерінде зерттеушілер іргелі мәселелердің бірі ретінде сәйкестік, аудармадағы заңды сәйкестік ұғымдарын атаған болатын [135]. Бұл түсініктің енгізілуі аударма ісін үдеріс ретінде танып, оған тілдік тұрғыдан талдау жасауға мүмкіндік берді. Ғалымдар аударма жасау барысында тіларалық сәйкестіктердің болатындығын және олардың белгілі бір заңдылықтарға бағынатындығын дәлелдеді. Алайда бұл терминге байланысты орыс тіл білімінде де, қазақ аударматану ғылымында да әлі де болса түйінді мәселелер баршылық. Біріншіден, әлемдік лингвистикадағы ғалымдардың сәйкестіктерді топтастыруға келгенде, тұжырымдары әртүрлі, ал қазақ мәдениетіндегі аудармаға арналған пікір-көзқарастарда нақтылық жоқ.

Зерттеуші Я. Рецкер сәйкестіктерді үш топқа ажыратады: а) балама (эквивалент); ә) ұқсастықтар (аналогии); б) баламалы ауыстырулар [136, 74]. Ал Л. Бархударов әр тілдегі сөздердің семантикасында елеулі айырмашылықтар болатындықтан, семантикалық сәйкестіктерді атайды да, олардың негізгі түрлері толық сәйкестік, ішінара сәйкестік және сәйкестіктердің болмауы деп ұйғарады [137, 30]. Басқа еңбектерден абсолютті сәйкестік, вариантты сәйкестіктер де болады деген ой ұшырасады.

Сәйкестік – түпнұсқа мәтіннің тілдік бірліктеріне аударылатын тілден таңдалынатын вариант; варианттардың бірі. Ол толық сәйкестік болуы мүмкін, бұл, әдетте, түпнұсқадағы сөзге, сөз тіркесіне екінші тілдің арнаулы сөздіктерінде берілген бастапқы мағынадағы сөз дегенді білдіреді. Екі тілдегі тілдік бірліктер мағынасы, функционалды қолданысы жағынан толық дәрежеде тұспа-тұс келеді. Бұндай тілдік бірліктер контексте тәуелді болмайды, жеке, контекстен тыс тұрғанда да мағыналары тұспа-тұс келеді. Мәселен, орыс тіліндегі понедельник; қазақ тіліндегі дүйсенбі, ағылшын тіліндегі Monday толық, абсолютті сәйкестік, эквиваленттер болып саналады. Л. Бархударов жер-су атауларын (мыс., Москва – Мәскеу), ғылыми-техникалық терминдерді толық сәйкестікке жатқызады. Я. Рецкердің эквивалент деп отырғаны да осыған саяды.

Екі тілдің лексикалық единицаларын салыстыру кезінде түпнұсқадағы сөздің лексикалық мағынасына аударылатын тілдегі сөздің бірнеше семантикасы сәйкес келген жағдайда ішінара сәйкестік пайда болады. Қазақ тіліндегі жүрдек деген сөздің орыс тілінде бірнеше мағынасы бар: *быстроходный, скорый, резвый*. Тәржіма ісінде түпнұсқадағы тілдік бірліктің қандай сөзбен тіркесімге түсіп тұрғандығын ескере отырып, осылардың бірін таңдау – контекстуалды сәйкестікті, функционалды сәйкестікті іріктеп алу деген сөз.

Сәйкестіктердің болмауы деген терминнің аясына екі тілдегі баламасыз лексика (мысалы, Африка елдерінің тілдерінде қар –с нег деген сөз жоқ), ұлттық реалий атаулары (мысалы, шаңырақ, сәукеле), кісі есімдері енгізіледі.

5.2 Қазақ және орыс тілдерінің параллель ішкорпусы: әзірлемесі, қолданылуы және мәселелері

Кез келген тілдің мәтіндер базасынан және оларға жасалған тілдік талдаулардан (белгіленімдерден) тұратын іздеу жүйесі бар ақпараттық ресурсы саналатын тілдік корпус жасау тәжірибесі ағылшын тілінің Лондон-Лунд (The London-Lund Corpus) ауызша тіл корпусынан (1960) басталып, Браун корпусы (The Brown Standard Corpus of American English) (1963), Ланкастер-Осло/Берген (LOB), Браун корпусының британдық ағылшын нұсқасы, Frown және FLOB, Brown және LOB баламалары және т.б. корпусар пайда болды.

Қазақ тілінің ұлттық корпусының (ҚТҰК) базасында параллель ішкорпустарды құрудың нәтижесінде қазақ тіліндегі мәтіндер мен басқа тілдердегі теңестірілген аударма мәтіндер базасын қалыптастыру арқылы тіл білімі (контрастивті лексикология, екітілді және көптілді лексикография, терминология, аударма теориясы, лингвомәдениеттану, когнитивті лингвистика), әдебиеттану (мәтіндердің аудармаларына текстологиялық талдау жасау), мәтінді автоматты тану (компьютерді мәшинелік аударма), әдістеме (қазақ тілін оқытудың лингвистикалық платформасын жасау, шет тілдерін оқыту) және т.б. салалардағы зерттеулерге тілдік материал жинақталады.

Қазақ және орыс тілдеріндегі параллель ішкорпусты әзірлеудің алғашқы сатысында екі тілдегі тілдік материалды электронды пішінге келтіру, электронды пішінге келтірілген мәтінді мазмұндық-семантикалық жағынан теңестіру (выравнивание), лингвистикалық белгіленімдер әзірлемесін жасауда және пайдаланушыға қажетті сөзді жылдам табуға мүмкіндік беретін «Іздеу» жүйесі интерфейсін жетілдіруде компьютерлік-бағдарламалық әдістер, сондай-ақ қазақ және орыс тілдеріндегі параллель ішкорпусты әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары, сондай-ақ client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.

Параллель ішкорпустар негізінен екітілді және көптілді болып бөлінгенімен, бұл ішкорпустың тілдік диалектілер салыстырылатын біртүрлі түрі де кездеседі. Екітілді және көптілді ішкорпустарды негізгі екі түрге бөлуге болады: 1) қандай да бір тілдің жазылған түпнұсқа мәтіндерінен және олардың бір немесе бірнеше тілге аударылған мәтіндерінен тұратын ішкорпустар; 2) екі немесе бірнеше тілде жазылғанына қарамастан, бір ғана тақырыптық саланың мәтіндерін біріктіретін ішкорпустар. Екітілді ішкорпустарға каталан-испан (<https://catalog.elra.info/en-us/repository/browse/ELRA-W0053/>), хорват ағылшын (<https://www.clarin.si/repository/xmlui/handle/11356/1058>), ағылшын-чех (<https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-1932>) сияқты ішкорпустарды жатқызуға болады. Аталған ішкорпустар бір-біріне жақын, туыстас тілдердің параллель мәтіндерінің негізінде жасалған. Ішкорпустардың базасына екі тілде шығарылатын журналдар, саяси партиялардың дебат материалдары, діни кітаптар мен көркем шығармалар және т.б. мәтіндер алынады. Әдетте параллель ішкорпустардың базасы корпусың мақсатына сай жинақталады. Ал көптілді параллель ішкорпус – ішкорпустың кең тараған түрі саналады. Олардың не-

гізгілерін атап өтсек: Виго университетінің лингвистикалық корпусы (CLUVI) (<http://sli.uvigo.es/CLUVI/index.php?corpus=21&tipo=8&lang=gl>), чех тілінің ұлттық корпусы (UČNK) (<https://trek.korpus.cz/>), орыс тілінің ұлттық корпусы (НКРЯ) (<https://ruscorpora.ru/new/search-para-en.html>). Бүгінгі күні аталған ішкорпустардан басқа да көптеген көптілді ішкорпустар бар.

«Параллель ішкорпус» ұғымына ғалымдар әртүрлі анықтамалар береді. Солардың бірнешеуін келтіре кетейік: «Параллельные корпусы являются своего рода сборниками стратегий и эквивалентов перевода, которыми руководствовались и которые придумывали переводчики» [45, 48]; «параллельный корпус можно определить как корпус, содержащий исходные тексты и их переводы» [138, 47]; «параллель корпустар – қайсыбір негізгі тілде жазылған түпнұсқа мәтіннің және осы негізгі мәтіннің бір немесе бірнеше басқа тілдерге аударма мәтіндерінің жиынтығының электрондық қоры» [12, 117]. Көріп отырғанымыздай, берілген анықтамалардан параллель ішкорпустың түпнұсқа мәтін және оның аудармаларының жиынтығы екенін көреміз.

Параллель ішкорпустарды әртүрлі лингвистикалық зерттеулерде (контрастивті лексикология, екітілді және көптілді лексикография, терминология, аударма теориясы, лингвомәдениеттану, когнитивті лингвистика), салғастырмалы әдебиеттану (мәтіндердің аудармаларына текстологиялық талдау), мәтінді автоматты тану (компьютерлік мәшинелік аударма), әдістеме (шет тілдерін оқыту) және т.б. салалардағы зерттеулерде тиімді қолдануға болады [139; 5].

Параллель ішкорпусты пайдаланушылардың сөздерді сөздік арқылы іздеп жатпай, мәнмәтіннен тауып, оның басқа сөздермен тіркесімін қарауына, тілдердің грамматикасын, семантикасын және лексикасын салыстырмалы зерттеуге мүмкіндік береді. Алайда ғалымдар параллель ішкорпустың кемшілігі ретінде мәтіндер көлемінің шағындығын және жанрлар бойынша шектеулілігін атайды. Сонымен қатар параллель ішкорпустың сапасына аудармашының қабілеті мен стилі ықпал етеді [140, 63-72].

Параллель ішкорпустардың негізгі түрлері

Орыс тілінің ұлттық корпусының параллель ішкорпустары мынадай тілдерді қамтиды: орыс, ағылшын, неміс, француз, испан, итальян, польша, украин, белорусс, армян, болгар, эстон және латыш [139, 383-401].

PARASOL корпусы (PARAllel corpus of Slavic and Other Languages) – Регенсбург университетінде жасалған параллель корпус [141, 123-138].

InterCorp корпусы – Чех ұлттық корпусының бір бөлігі (www.korpus.cz/intercorp).

PARUS корпусы (PARalelní RUsko-Slovenský korpus) – Штур атындағы Тіл білімі институты жасаған параллель корпус (<http://korpus.sk:8099/manatee.ks/index>).

ASPAC корпусы – Амстердам университеті жасаған көркем мәтіндердің және олардың аудармаларының көптілді корпусы (ASPAC – Amsterdam Slavic Parallel Aligned Corpus) (<http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/b/a/a.a.barentsen/a.a.barentsen.html>).

Параллель ішкорпустар метаақпараттан (әрбір мәтін туралы ақпарат: қашан жасалды, кім жасады, көлемі және т.б.) және белгіленімнен (әрбір сөздің лингвистикалық «төлкұжаты») тұрады.

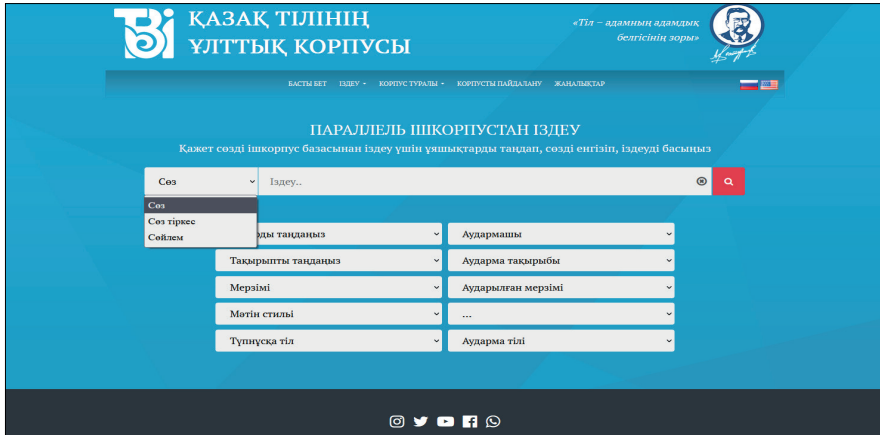
Параллель ішкорпустардағы белгіленімнің маңызды элементі – түпнұсқа тілдегі әрбір тарауға, абзацқа, сөйлемге аударма тілдегі тараудың, абзацтың, сөйлемнің сәйкес келуін теңестіру. Осылардың ішінде сөйлем бойынша теңестіру тиімдірек әрі пайдалырақ болады. Бұл жұмыс қиын да ұзаққа созылады әрі қымбатқа түседі, сондықтан әлемде сөйлем деңгейінде теңестірумен жасалған параллель корпустар саны аз.

Кейбір елдердің, мысалы, Орыс тілінің ұлттық корпусында параллель ішкорпустардағы мәтіндерді теңестіру үшін арнаулы бағдарлама жасалған. Алайда мұндай бағдарлама арқылы теңестірудің техникалық тұрғыдан қиындығы да бар. Автоматты түрде теңестіру сөйлем деңгейінде, таза құрылымдық (сөздердің, сөйлемдердің ұзындығы, саны) және статистикалық (сөйлемді құрайтын сөздердің жиілігі) әдістермен жасалатындықтан, теңестіру процесі қайта тексеріліп, «қолмен өңдеу» жүргізіледі. Мысалы, ағылшын және неміс тілдеріндегі корпусты жасаушылар (А.А.Кретов, И.Е.Воронина, Д.Спесивцев) қолданатын «ПарТекс», сондай-ақ басқа параллель корпустарды теңестіру үшін «Евклид» (Д.В.Сичинава, Т.А.Архангельский) бағдарламасы бойынша теңестіру сөйлем деңгейінде жасалады. Теңестірудің айтарлықтай қиындығының бірі – автордың мәтінді сөйлемге және абзацқа бөлудің аудармашының бөлуімен сәйкес келе бермеуі. [141, 194-234]. Яғни аудармашы сөйлем шегін ешқашан сақтамайды десе де болады, түпнұсқадағы сөйлем екі немесе бірнеше сөйлеммен және керісінше аударылуы мүмкін.

ҚТҰК-ның қазақ мәтіндері мен олардың орыс тіліне аударылған нұсқаларының параллель ішкорпусын жасау үшін алғашқы тәжірибе ретінде М.Әуезовтің «Абай жолы» роман-эпопеясы және оның А.Ким аударған орыс тіліндегі мәтіні алынды.

ҚТҰК-ның параллель ішкорпусының сипаты

Әзірленіп жатқан параллель ішкорпус теңестірілген мәтіндер базасынан және іздеу жүйесінен, ал енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл.* (Автор: М.Әуезов, Аудармашы: А.Ким, Шығарма атауы: Абай жолы, Аударма атауы: Путь Абая, Мерзімі: 1942, 1947, 1952, 1956, 2009, 2013 жыл, Аударылған мерзімі: 2012 жыл, Түпнұсқа тіл: Қазақ, Аударма тіл: Орыс). Іздеу жүйесі қандай да бір сөзді сөз, сөз тіркесі, *сөйлем, бас әріп* (қазақ және орыс тіліндегі) параметрлері арқылы табуға мүмкіндік береді.



Сурет 20. Параллель ішкорпустаң интерфейсі

20-суретте қазақ және орыс тіліндегі параллель ішкорпустаң интерфейсі көрсетілді. Парақшада көрсетілген іздеу параметрлерінің бірі арқылы қажетті сөзді іздегеннен кейін нәтижелер интерфейсі ашылады.

Табалған құжаттар саны «бала» - 64	
<p>Мұхтар Әуезов - «Абай жолы», 1-том</p> <p>Үш күнді жолдан бүгінгі, соңғы күніне шаққар баға баран салды. Қорықтан күш шыға аға мінбейсі деп ақыл қалды. Пұты қалаған алып қайтала барған ағылған Байғасы да тақ атар-ағылған оқ өткен тұрғыны сай. Құрғаны аттан да түсей, көзе жұртшыларын оқ бойға аға отқанды. Көй-жөке өлке талас. Көкқарым мен Бұратын, Тақырұлақ сияқты қолма-құданқарды түс-тұрғын көпесте баян оқышпай, астындағы қаруы құла бөстін ағыл-ағылға та алды.</p> <p>Мұхтар Әуезов - «Абай жолы», 1-том</p> <p>Арттылар - Байғас пен жорға Жұрабай: - Мына баламыз аулетке асығып-ай! Сөзге баға қастай іш-құста болып қалған-ау, - деді. Шаңқар баға ұзын кеткенде бұлар да ерсіге жеке шоқандары қауаты отқанды көзі көетей. Жорға Жұрабайдан тағымына қара шоқанды бар. Байғастың да ағылған басына тігер ұзын айыры соғыла бар-ды. Тақырұлақ тұрғын көпесте Байғастың бағаны жеке шабуыл текен:</p> <p>Мұхтар Әуезов - «Абай жолы», 1-том</p> <p>-Өңіреті сендердің армен жоқ екен, еңді еңдіре жұрдым не, жеке жұрдым не? Ақ көтпейі - деп, шоқанды жіберей. Бұл - Тақырұлақтан өте берісте бағатым ай. Саған жағаны қауаты деген Есембайға жеткенде, алдына бірде-бір қараны жоқ. Көз ұшына көгіз ұзын алып, қолыңды жағалы қауаты отқанды. Жаман бұл тұттары мін беке-беке-беке, асы қауатындай боп жүр Шығайға, жағалы қарай көнетен де етсі боп қалатын жүр. Ағылған жолды баян отқанды тұрғыны бар. Тақ неқ артқан, өлке тұрғын жұртшыға мақал қолы-қолты, құшақтары түсіретін ұры сай, жасырын ағарыды да бар. Бұдан бұрынғы екі күнді жолға үнемір жылдан жүрмей, бағанын шиданды ойын тұрғын-ды. Оң соңдақтан, бүгін өзгені жеткен күні үнемірді ерсісі қатты жүретінді айығын тағымына деп жорға. Күні бойы осылай еуге байтаған.</p>	<p>Малына спеңді домой. Он отғор был на все, чтобы третий день пути был и последний. На почкема в Корым (1 - В казахском языке упреждение ставится на последний слог: Корым, Байғас и т.д.) он затенно разбудил Байғаса - родственника, проезжающего за ним в город, и угостил своих спутников выехать, едва занявась зари.</p> <p>Весь день он подгонял коня, держась впередии провожаемых на расстоянии пушистой стрелы, Байғас и старый Жұрабай только восклицали: - Ну и топчется же вылетает в путь! - Велелай. Видно, всея заны прощала в мерее? - Мерее - мусульманское училище.] со скуки! И оба шевелили коней, то раслы, то вскачь догоняя его. - Жұрабай - жаманша оқ қолмамы шыр қарай шоқар, Ш - Шоқар - дүдина с ұтопешем на қолше, оғарың рұқшашың боп. Байғас - приделывающа миссон сапоз дивный ореховой соды.]- Сопы - динная, ұтопешем в коню палы, ұтопешем в коню боп.]) Воле урочна Тақырұлақ Байғас, умеря лия подостра, крикну ему:</p> <p>- Ну, раз и вы ничего не сможете сделать, так не все ли равно, абы куда или в каком? И повел. - Он затен коня, поспал вперед и до самой пошине Есембай ни разу не отступил назад. Два перама деп пути старшие двигались не спеша и соковы вывети малына из топешем. Телер он был рад, что жолы в последний день вывет средство закончить их топешем; он решил до самого зана сказать вперед. Спутники почти потеряли его из виду, но он все скавал. Путь пролегал по буртам и сопкам. Когда ауыл откопываеат в горы Чингиз, здесь становите совсем безлюдно. С каждого холма можно отчетно слышать за проселками, и пологими до чужого добра негдетно выаждат на все из отарго и пошн.</p> <p>The boy was in a hurry to go home. He was ready to do anything to make the third day of the journey the last. At an overnight stay in Koruk (1 - in the Kazakh language, the stress is on the last syllable: Koruk, Baytas, etc.), in the dark, he woke Baytas, a relative who came to the city for him, and persuaded his companions to leave as soon as dawn broke.</p> <p>All day he urged the horse, keeping ahead of the guides at the distance of an arrow shaft, Baytas and old Zhurabay only exclaimed: - Well, the little boy is in a hurry to go to the aull! - Poor fellow! - Apparently, all winter disappeared in the madrasah (1 - Madrasah - Muslim school.) Out of boredom And both moved their horses, now at a trot, now at a gallop, catching up with him. Zhurabay - clatching his mule's shokpar under his knee, (3 - Shokpar - a club with a thickening at the end, a hand-to-hand combat weapon.) Baytas - holding a long birch snail with the toe of his boot. (4 - Snail is a long stick, thickened towards the end, used in equestrian combat.) (Near the Taktikubul tract, Baytas, tempering the ardor of a teenager, shouted to him:</p> <p>- Well, since you can't do anything either, does it really matter if I'm alone or with you? I went! - He hit the horse, galloped forward and never looked back to the very hollow of Yeserbul. The first two days of the journey, the elders moved slowly and completely put the boy out of patience. Now he was glad that at least on the last day he had found a means to finish their hurry: he decided to gallop ahead to the village itself. The companions almost lost sight of him, but he kept galloping. The path ran along hillsides and hills. When the aulls migrate to the Chingiz mountains, it becomes completely deserted here. From each hill you can perfectly follow the passerby, and it is not difficult for hunters to the good of others to attack them from ravines and hollows.</p>

Сурет 21. Параллель ішкорпустан сөз іздеу нәтижесінің интерфейсі

21-суретте параллель ішкорпустан «бала» сөзін іздегеннен кейінгі іздеу нәтижесі көрсетілген. Мұнда кестенің сол жағында түпнұсқа, ал оң жағында аударма мәтін орналасқан. Ішкорпустан іздеген «бала» сөзі көк бояумен ерекшеленген. Нәтижелер интерфейсіндегі мәтіндерді «көшіріп алу» функциясы жұмыс істейді.

М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы параллель ішкорпустан абзац (сөйлем) денгейінде қолмен теңестірілді. Мұндай қолмен теңестіру жұмыс күшін және уақытты талап еткенімен, кейін параллель ішкорпустаң базасын жинақтап, өңдеуге арналған жасанды интеллекттің үйретуге негіз болып табылады.

Бұл параллель ішкорпустаң әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары қолданылды.

Паралель ішкорпустың бағдарламалық архитектурасын және жұмыс істеу тәртібін төмендегіше көрсетуге болады (22-сурет):

1) екі тілдегі мәтіндер Excel офистік бағдарламасы арқылы жинақталып, абзац (сөйлем) деңгейінде қолмен теңестірілді;

2) теңестірілген мәтіндер Excel файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктелді;

3) жүктелген мәтіндер Notepad++ мәтіндік процессор бағдарламасы арқылы сұрыпталып, олардың статистикасы алынды;

4) сұрыпталған мәтіндерді интернетке жариялау және пайдаланушының сұрауларын қамтамасыз ету үшін Django веб сервері қолданылды;

5) Django веб серверін MySQL мәліметтер қорын басқару жүйесімен байланыстыру үшін Python тілінде жазылған, іздеу функциясымен қамтамасыз етілген Processing.py бағдарламасы қолданылды;

6) параллель ішкорпустың бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.



Сурет 22. Паралель ішкорпустың бағдарламалық архитектурасы

Параллель ішкорпус бойынша статистикалық мәлімет

Параллель ішкорпусқа енгізілген «Абай жолы» роман-эпопеясының мәтінде қолданылған жалпы сөз саны: қазақ тілінде – 482150, орыс тілінде – 445491; қайталанбайтын сөз саны: қазақ тілінде – 59816, орыс тілінде – 59952, сөйлем саны: қазақ тілінде – 41798, орыс тілінде – 37792 екені анықталды. Екі тілдегі мәтінде ең жиі қолданылған сөзқолданыстың ондығына Абай сөзі кіреді: қазақ тіліндегі мәтінде қолданылу жиілігі бойынша 4-орында, орыс тіліндегі мәтінде – 10-орында.

Кесте 7

«Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтнінде және орыс тіліндегі аударма мәтнінде жиі қолданылған он сөзқолданыстың сандық және пайыздық көрсеткіші

№	Қазақ тіліндегі мәтінде			Аударма мәтінде		
	Сөз	Кездесу жиілігі	%	Сөз	Кездесу жиілігі	%
1.	да	5682	1,18	и	18356	4,12
2.	деп	4554	0,94	в	10769	2,42
3.	де	4487	0,93	не	8408	1,89
4.	Абай	4091	0,85	с	7907	1,78

5.	бір	3937	0,82	что	5882	1,32
6.	осы	3490	0,72	он	4530	1,02
7.	мен	3147	0,65	его	4065	0,91
8.	бар	3039	0,63	к	3694	0,83
9.	бұл	2872	0,60	по	3298	0,74
10.	жоқ	2417	0,50	Абай	2820	0,63

Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтінін орысша аударма нұсқасымен теңестіру барысында мынадай ерекшеліктер байқалды:

1. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) тұрпат межесі тұрғысынан, яғни абзацта (сөйлемде) қолданылған сөздер сан жағынан шамамен сәйкес келеді. Мысалы:

1) – *Ойбай, бізде не қауқар бар? Біз екеу-ақ...* ± – Ойбой! А что мы сможем с ними сделать? Нас двое...

2) – *Әй, балам-ау, мына жер – жау жатағы. Бұл ұрының ойнағына кеп алып, жаман ырым бастағаның не қылғаның? – деді.* ± – Ой, сынок, накличешь ты беду своими шутками! Нашел место – в самом воровском логове, – сказал он с досадой.

3) – *Е, тастай алмай жүрген жаным жоқ! Терге демесем, жанымды ала зой деп кепілге келіп пе екемін, – деп томсарып қалды.* ± – У меня тоже нет души, которую некуда бы девать! Я сказал лишь, что надо проверить. Не для того я приехал, чтобы заложить свою душу, – проговорил он мрачно.

2. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) мазмұн межесі тұрғысынан шамамен сәйкес келеді. Мысалы:

1) Баланың қытығына тиетіні осы жері. ± – *Но эти слова только подзадорили мальчика.*

2) Содан жаңағы қауіпті деген Есембайға жеткенше, артына бірде-бір қараған жоқ. Көз ұшына кетіп ұзап алып, ылғи жапа-жалғыз шауып отырған. ± – *Он ударил коня, поскакал вперед и до самой лощины Есембая ни разу не оглянулся назад.*

3) – *Одан да, не күтсеңдер де осы жаман қарадан күтсеңдерші! – деді.* ± – *Если уж чего-нибудь ждать – так ждите только от этого черномазого мальчугана!*

4) Кірпігін де сирек қағады. ± Он даже моргает редко.

5) – *Ендеше, өлімнен ұят күшті. Ел көрмеген сұмдыққа, ел көрмеген жаза керек! – деп Құнанбай байлауын айтты.* ± – Честь – выше смерти. Беспремерный грех должен получить и беспремерное возмездие, – заключил Кунанбай.

3. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аударма мәтіндегі абзацтар (сөйлемдер) тұрпат және мазмұн межесі тұрғысынан сәйкес келмейді: қазақ тіліндегі түпнұсқа мәтіндегі кейбір абзацтар (сөйлемдер) орыс тіліне бұрыс немесе үстүрт, қысқа аударылған, олардың жуық мағынасы берілген. Мысалы:

1) Үй іші Абай айтқан жаңалықтарды қызыға тыңдап, көңілденіп қап еді. ≠ Старшие чувствовали себя неловко: они были смущены своим незнанием.

2) Бұл үлкендер осы түнде қырық-елудей жігіт-желеңді және бірер ақылшы кәріні қастарына алып, өліктің артын күтетін қамға кірісті. ≠ – Они созвали на совет еще несколько стариков.

3) Оспан мен Смағұл Тәкежан қасына қатар жалынып, жүгіріп барып еді. ≠ – *Оспан и Смагул продолжали приставать к нему.*

4) Жұрттың бәрі амалсыз иіріліп тұра қалып, ақ таймен Оспанның арпалыс майданына қарады. ≠ Все невольно залюбовались необыкновенным единоборством.

5) Абай қып-қызыл боп кетіп, шұғыл ойланып отырып қалды. ≠ – *Абай вспыхнул.*

6) Байдалы мына сөзге әбден қанағат қылды. Бала көргені – бала емес тәрізді. ≠ Его слова понравились Байдалы.

7) «Ақ етің аппақ екен атқан таңдай!» дегеннен басқасы есіне әзір қайта түспеді. ≠ В памяти живет только одна строка: Ты встаешь а моем сердце, расцвет любви.

8) Шолпысының ендігі сылдырынан өзі қысылған тәрізді. ≠ Казалось, что каждый шаг ее звенит серебряными переливами дорогого украшения.

9) Ырғызбай ішінен, әсіресе, осы Қарашоқыға қызығушылар көп болатын. Бұнда отырған ауылдар Борсақ, Жексен ауылы. Ол, Борсаққа әлдеқандай көрінгенмен, өзге жұрттың келесінде бұта құрым болатын. ≠ У многих из рода Иргизбай давно уже глаза разгорались на Карашоку, где находился аул Жексена.

10) Абай да, Жұмабай да қызығып тыңдасты. Өзі сері, әнші, сұлу Байтасқа екеуі де – бірі кәрі, бірі жас бала – қызғана да, тамашалай да қараған еді. ≠ Байтас был признанным певцом и красавцем. Абай посмотрел на него с нескрываемым восхищением.

11) Үлкендер бала мінезіне бір күліп, бір таңданып сүйсініп қалды. Қара сұр бала көптің көзі бір өзіне қадалғанда енді қысылыңқырап, қызарып еді. Бірақ, көзінде жайнай жанған от білінеді. Өзге балаларының ажарынан бөлек, қызулы, саналы оты бар сияқты. ≠ Шутка мальчугана вызвала и смех и восхищение взрослых.

12) – *Бәсе, Шаншар?* ≠ – Ну, конечно, в нем течет кровь Шаншар.

Мұндай мысалдарды көптеп келтіруге болады. Алайда мұндай «ерекшеліктер» бір тілдегі көркем әдеби туындыны құрылымы өзгеше екінші бір тілге аудару кезінде қолданылатын оңайлату (аударма мәтін орташа алғанда түпнұсқаға қарағанда лексикасы және синтаксисі жағынан «қарапайымдандырылады»), нормаландыру (аудармашы аударма тілдің нормаларына сай тілдік бірліктерді қолданумен қатар шығыс және аударылатын тілдің иесі халықтың мәдени ерекшеліктеріне байланысты «өзгерістер енгізеді», «трансформация жасайды»), симпликация (лексикалық (аз сөз қолдану), синтаксистік (сөйлемнің синтаксистік құрылымын оңайлату) және стилистикалық (фразеологиялық оралымдарды жай сөз тіркестерімен беру және т.б.; түпнұсқа мәтіндегі ұзын және күрделі сөйлемдер аударма тілде қысқа және жай сөйлемдерге бөлінеді) [142, 18-27], эксплицитация (кеңейту (addition) – аударма мәтінге жаңа элементтер қосу (сипаттау, күшейту және т.б.) және нақтылау (specification) – түпнұсқа мәтіндегі имплицитті ақпараттың аударма мәтінде эксплицитті түрде берілуі, бұл түпнұсқа мәтінмен

салыстырғанда аударма мәтіннің көлемінің ұлғаюына алып келеді) [143, 36-47] тәсілдерінің бірін қолдануға және т.б. себептерге байланысты болады.

М.Әуезовтің «Абай жолы» роман-эпопеясының орыс тіліне аударылу мәселелерін әдебиеттану теориясы тұрғысынан зерттеген ғалымдардың қатарынан Г.Бельгер, М.М.Әуезов, Қ.Алпысбаев, А.Анастасьев, С.Ананьева, А.Жақсылықов, Н.Сағындықова, Г.Қазақбаева, Г.Қазыбек, Ж.Каракузова, Б.Канапьянов, С.Елеуқенов, Г.Болатова және т.б. атауға болады.

Көрнекті аудармашы Г.Бельгер «Абай жолы» роман-эпопеясының орыс тіліндегі аудармаларына (А.Никольская, Л.Соболев, Т.Нуртазин, Я.Смеляков, В.Бугаевский, А.Жовтис, З.Кедрина) талдау жасай келе, А.Кимнің аудармасына қатысты: мәтіннің орыс тіліне кейде өлең жолдарымен аударылған тұстары байқалады; қазақтың сөзсаптамына тән интонация, салыстыру, мақал-мәтелдерді, ұйқасты сөз орамдарын және т.б. жақсы түсініп, көбінесе дәл бере алған; аудармада солғын көпсөзділікке ұрынған тұстар кездеседі; қазақ халқының шешендік сөз өнеріне қатысты қолданыстар көбінесе «краснобай» («сөзуар») сөзімен аударылған; кейде аудармада орыс тілінде орашалақ, ыңғайсыз, өңсіз «сипаттағы» сөздер қолданылған; кейбір сөздер барабар баламасымен аударылмаған; аудармада тұрмыстық, этнографиялық сипаттағы «ағаттықтар» кездеседі; ономастикалық атаулардың транслитерациялануында сәйкессіздіктер кездеседі (Үдері – Одели, Барлыбай – Барылбай, Алмағамбет – Алмагамет, Тайлақбай – Гайлақбай, Жаманқарин – Жаман-Гарин, Жаман-гарин; Семей – Семипалатинск); түпнұсқада берілмеген жекелеген біртегіс құрылымдар көзге түседі; кейбір сөздер дұрыс аударылмаған (томаға – подстава); Мағрипаның ішкі көңілі сипатталатын бір бет аударылмай кеткен [144, 130-142] деген пікір айта отырып, мынадай тұжырым жасайды: «Надеюсь, новый перевод «Пути Абая», осуществленный А.Кимом, вдохновит переводчиков этого творения на иностранные языки» [144, 143].

Ғалым А.Жақсылықов роман-эпопеяның А.Ким жасаған аудармасы туралы: «стиль толық, экспрессивті бояуы жақсы берілген» деп баға бере келе, мынадай кемшіліктерді топтап көрсетеді: «а) аудармашы қазақ халқының ұлттық менталитетін өрнектейтін этномәдени және этнопсихологиялық алғышарттарды ескермейді (мысалы, жылқының жасына, түсіне, жынысына қатысты сөздер); ә) қазақ ру-тайпаларының шежіресін, құрылымы мен қатынастарын шатастырып алады, б) аудармашы ХІХ ғасырдағы қазақ қыздарының лексикасында модернизмдерді (сленглер мен урбанизмдерді) қолданады [145, 167-178].

Зерттеуші Г.Болатова М.Әуезовтің «Абай жолы» роман-эпопеясының А.Ким жасаған аудармасына қатысты зерттеулерге шолу жасай келе: «М.О.Әуезовтің роман-эпопеясын орыс тілінде толығымен қайта жаңғырту мәселесі әлі де болса түбегейлі шешімін тапқан жоқ. Осылайша, роман-эпопеяның орыс тіліндегі аудармалары әлі күнге сапа жағынан жетілдіру қажеттігін талап етеді. Демек, оның басқа тілдердегі аудармалары да толыққанды талдау жасауды қажет етеді, өйткені басқа тілдерге орыс тілінен аударылғаны белгілі», - деген тұжырым жасайды [146, 323].

Осы орайда «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен орыс тіліндегі аудармасын теңестіру барысында төмендегідей жайттардың да байқалғанын атап өткізіміз келеді:

1. Роман-эпопеяның қазақ тіліндегі түпнұсқа мәтіндегі лексикалық бірліктер орыс тіліне қате аударылған. Мысалы:

1) Өзі Абайдың 16 жасында туып ерте ересек боп қалған соң және ата қолында оңаша үйде жалғыз өскен кенже есепті болған соң. ≠ Так как Абай был старше своего первенца всего на семнадцать лет, и тот рос с младенческого возраста в доме деда, Акылбай не признавал отца и не испытывал к нему сыновних чувств.

2) Қазір жасқа толған, нұрлы көздері көп уақыт кірпік қақпай, жігітке ұзақ тұнжырап, қарап қалды. ≠ Она отошла к окну и долго вглядывалась в темноту наступившей ночи глазами, полными слез...

2. Роман-эпопеяның орыс тіліндегі аудармасында сөз ортасындағы қысаң «ы» әрпі түсіріліп жазылған. Мысалы, ауыл (қаз) – аул (орыс.), атрап, баур (қаз) және т.б.

3. Роман-эпопеяның орыс тіліндегі аудармасында топонимдерді, этнонимдерді, антропонимдерді транслитерациялауда әртүрлі сәйкессіздіктер кездеседі: емле кателері, ру атаулары қазақ тіліндегі түпнұсқа мәтінде бас әріппен, орыс тіліндегі аударма мәтінде кіші әріппен беріледі, бір ғана атау орыс тіліндегі аударма мәтінде бірнеше түрде жазылады. Мысалы, Керімбала – Коримбала, Есембай – Осембай, Көжекбай – Кожекбай, Ойке – Ойкал, Керей – керей, Уақ – уаки, Қорық – Корык, Корык және т.б.

Осы ретте академик-жазушы М.Әуезовтің әлемдік деңгейдегі «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі аударма мәтінін параллель ішкорпусқа енгізу барысында қолданылған теңестіру әдісі арқылы негізінен екі тілдегі мәтіннің абзац (сөйлем) деңгейінде мазмұндық және құрылымдық тұрғыдан сәйкес келуі басшылыққа алынғандықтан, роман-эпопеяның аудармасының сапасын анықтау, яғни лингвистикалық (стилистикалық, орфографиялық және т.б.), әдеби (көркемдік) және т.б. талдау мақсаты қойылмағанын атап өтеміз.

ҚТҰК құрамында әзірленіп жатқан М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы параллель ішкорпуста абзац (сөйлем) деңгейінде қолмен теңестірілді. Сонымен қатар параллель ішкорпусқа Ж.Аймауытовтың «Ақбілек», Т.Ахтанов «Боран», С.Мұқановтың «Мөлдір махаббат», Ш.Айтматовтың «Боранды бекет», Ғ.Мүсіреповтің «Ұлпан», С.Сейфуллиннің «Тар жол тайғақ кешу», Т.Ахтановтың «Боран», Ә.Кекілбаевтың «Аңыздың ақыры», Ә.Нұрпейісовтің «Соңғы парыз» романының, С.Елубаевтың «Ақ боз үй» роман-трилогиясының, С.Мұратбековтің «Жусан исі» повесінің, М.Жұмабаевтың «Шолпанның күнәсі», М.Әуезовтің «Қорғансыздың күні», «Көксерек», М.Мағауиннің «Бір атаның балалары», О.Бөкейдің «Жасын, Ардақ, Бура», Д.Исабековтің «Тыныштық күзетшісі», А.Кемелбекованың «Қоңырқаз», Ж.Шаштайұлының «Есбай тегі», Р.Мұқанованың «Мәңгілік бала бейне», Ш.Бейсенованың «Мизамшуақ» әңгімесінің және т.б. көркем шығармалардың және ісқағаздарының қазақ тіліндегі түпнұсқа мәтіндері мен орыс тіліндегі аударма мәтіндерін теңестіру қолға алынып отыр.

Параллель ішкорпустың бағдарламалық архитектурасын және жұмыс істеу тәртібін төмендегіше көрсетуге болады: 1) екі тілдегі мәтіндер Excel

офистік бағдарламасы арқылы жинақталып, абзац (сөйлем) деңгейінде қолмен теңестірілді; 2) теңестірілген мәтіндер Excel файлынан тікелей MySQL мәліметтер қорын басқару жүйесіне жүктелді; 3) жүктелген мәтіндер Notepad++ мәтіндік процессор бағдарламасы арқылы сұрыпталып, олардың статистикасы алынды; 4) сұрыпталған мәтіндерді интернетке жариялау және пайдаланушының сұрауларын қамтамасыз ету үшін Django веб сервері қолданылды; 5) Django веб серверін MySQL мәліметтер қорын басқару жүйесімен байланыстыру үшін Python тілінде жазылған, іздеу функциясымен қамтамасыз етілген Processing.py бағдарламасы қолданылды; 6) параллель ішкорпустағы бағдарламалық архитектурасы client-server және MVC (Model-View-Controller) технологиялары қолданыла отырып әзірленді.

Қазақ және тілдерінің орыс параллель ішкорпусы теңестірілген мәтіндер базасынан, белгіленім, метабелгіленім және іздеу жүйесінен және ішкорпуста енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл*. Іздеу жүйесі қандай да бір сөзді сөз, сөз тіркес, сөйлем, бас әріп (қазақ және орыс тіліндегі) параметрлері арқылы табуға мүмкіндік береді.

Әзірленген қазақ және орыс тілдерінің параллель ішкорпусы қазақ тілі мен өзге тілдердің теңестірілген мәтіндері арқылы қазақ тілінің қолданысын кеңейтетін цифрлық платформаның бір үлгісі, тіл білімі, әдебиеттану, әдістеме және т.б. салалардағы зерттеулерге қажетті мәтіндер базасы бола алады. Сонымен қатар бұл параллель ішкорпус қазіргі кезде көптеген тілдердегі түрлі стильдегі мәтіндерді қамтитын <https://opus.nlpl.eu/> және т.б. платформалар сияқты түркі тілдерінің мәтіндерін салыстыруға арналған осындай платформаларды, яғни параллель ішкорпустарды жасаудың бастамы деп санаймыз.

5.3 М.Әуезовтің «Абай жолы» роман-эпопеясының орысша аудармасындағы бейвербалды амалдар

Жарық көргеніне 80 жылдан асқан қазақ әдебиетінің үздік туындысы «Абай жолы» роман-эпопеясы тілдік тұрғыдан түрлі құралдарға, соның ішінде бейвербалды амалдарға аса бай. «Абай жолы» роман-эпопеясы әлемнің 116 тіліне аударылып, 200 томдық «Дүниежүзілік әдебиет кітапханасына» енді және XX ғасырдың ең үздік шығармаларының бірі болып саналады.

«Абай жолы» романын тілдік тұрғыдан алғаш зерттегендердің бірі Е.Жанпейісов «*Абай жолы*» – *әдебиеттің ғана емес, аса бағалы тіл ескерткіші де болып табылады. Онда қазақ тілінің байлығы, сарқылмас мүмкіншіліктері, ғасырлар бойына қалыптасқан сөздік қорының арғы-бергі қат-қабаты барынша мол қамтылған. Төрт кітаптан тұратын бұл шығармада осы күнгі сөздіктерде кездесе бермейтін де тіл құбылыстары жиі ұшырайды... Әр тарауы этнографиялық жағынан бас-аяғы бүтін жеке-жеке ғылыми еңбек десе де болады*» [147, 8] деп баға береді.

Қазақ халқының ұлттық дүниетанымының бір бөлігі саналатын бейвербалды амалдардың тілдегі көрінісін, вербалдану ерекшеліктерін қарастыруда

«Абай жолы» романының құндылығы жоғары деп есептейміз. Бұдан бөлек, қазақ халқына тән бейвербалды амалдардың басқа бір мәдениетке берілуін, яғни, аударылу деңгейіне талдау жасау үшін де бұл романның деректік құндылығы жоғары. Шығармадағы паралингвистикалық құралдардың аудармасы шығарма мазмұнының толық ашылуына, оқырман қабылдауы мен бағалауына тікелей әсер етеді.

Бейвербалды амалдардың вербалдануы деп коммуникация кезінде қосымша ақпарат беретін ым-ишарат, дене қозғалысы, дауыстағы өзгерістер сынды тілдік емес құралдардың мәтіндегі көрінісін айтамыз. Тіл білімінде бейвербалды амалдарды қарастыратын антропоэлекті бағыттағы сала *паралингвистика* деп аталады. Профессор Б.Момынова паралингвистикаға: «Паралингвистиканың лингвистикалық жүйеде өз орны бар, көмекшілік қызмет атқаратын, сонымен қатар тілдік емес құралдардың нақты бір сөйленіс кезіндегі жұмсалымдық қызметін, мағыналық жақтан түрленімін зерттейтін, функционалды сала» [148, 34] – деген анықтама береді. Олай болса, бейвербалды амалдар дегеніміз:

- коммуникацияның көмекші құралы;
- көркем шығарма тіліне эксперссия беруші негізгі элемент;
- кез келген ұлттың танымын жеткізе отырып, әлемнің тілдік бейнесін жасаудың құралы;
- адамның түрлі көңіл күйі мен эмоционалды жағдайын білдіретін парамтоивті элементтер.

Көркем шығарма тіліндегі бейвербалды амалдарды қарастыруда оларды параамалдарға ұқсас, бірақ коммуникацияға қатыспайтын қимыл-қозғалыстардан ажыратып алу керек. Ақпарат беру функциясын атқармайтын, сол себепті бейвербалды құралдарға жатпайтын тілдік емес амалдардың үш тобы бар.

Біріншіден, «жеке адамға тән физиологиялық ерекшеліктер бейвербалды амалдарға жатпайды. Мұндай дағдыны тілді тұтынушы қоғам тұтас қолданбайды да, сондықтан параэлементтердің санатына енбейді, адамның физикалық күйіне жатқызылады және паратілдік жүйеге қойылатын өлшем-белгілерге, талаптарға жауап бермейді» [148, 19].

Екіншіден, ешқандай қосымша ақпараты жоқ жай ғана «қарау», «күлу», «тұру» секілді қимыл-әрекеттерді атаушы сөздер бейвербалды амалдарға жатпайды.

Үшіншіден, адамның ішкі күйін білдіретін, алайда оның сыртқы көрінісі мәтінде берілмейтін тілдік бірліктерді де бейвербалды амалдардың вербалданған түрінің қатарында қарастырмаймыз. Мысалы, ұялу, ренжу, ашулану т.б.

Әр ғылым саласы бейвербалды амалды өз қырынан, өзіндік концепциясы тұрғысынан зерттейді. Сол себепті барлығына ортақ бір ғана теория немесе бір ғана классификация жоқ. Сонымен қатар бейвербалды амалдар әртүрлі халықтар мен мәдениеттер үшін де түрлі мәнге ие. Бұл аталған жүйенің қаншалықты күрделі және заманауи көзқарастар тұрғысынан зерттеле түсуі қажет сала екенін көрсетеді. «Абай жолы» романында тілдік емес құралдардың мынандай төрт типі көп қолданылған [149]:

Просодикалық амалдар дауыстағы түрлі өзгерістер арқылы қосымша ақпарат береді: *біраз тұрды, аз тоқтап, қатты зекіп, қатты сөйлеп, сыбырлай күбірлеп, ашық дауыспен, дауыстан алды, күбір етіп, күңк етіп, қатты әмір етіп, зіркілдеп сөйлеп, ұрыса сөйледі, ақырын дауыспен, қатты ақырып, арыс-арс етіп, сыбырмен, сыбырлап, саңқ етіп, қатты ұрсып* т.б.

Проксемикалық амалдар коммуниканттардың кеңістіктегі орнына қарай прагматиканы көрсетеді: *қатар тұр екен, дәл қасына, төрге, от басына, аттылардың қалың ортасында, екі жағында, алға салып, оңды-солын, ортасында, артынан* т.б.

Кинесикалық амалдар сөйлеушінің бет-жүзіндегі, дене-қалыбындағы түрлі өзгерістер арқылы берілетін қосымша ақпаратқа негізделеді: *қабағын бір шытып, қозғалақтап қалды, ырғалып қалысты, қолын сілкті, қадалып қалды, тепсініп* т.б.

Такесикалық амалдар коммуниканттардың бір-біріне жанасуы арқылы берілетін түрлі бейвербалды кодтарды қамтиды: *мойнынан құшақтай алып бетінен сүйді, қойнына жабыса түсіп, кеудеден қойып қап, жұлқып қалып етпетінен салып, маңдайынан ұзақ иіскеп, маңдайына алақанын қойып, қысып, бауырына басып, арқаға қағып, жақтан тартып-тартып қалды, қойып-қойып жіберді, бетінен, мұрнынан қайта-қайта сүйді* т.б.

Бейвербалды амалдар көркем мәтінде вербалданған түрде көрініс табады. Алайда көркем шығарманы басқа тілге аудару кезінде параэлементтерді аудару айтарлықтай қиындық туғызады. Аударматанудың көрнекті өкілі Л.С. Бархударов: «Аударма дегеніміз – бір тілдің сөйлеу (тілдік емес) шығармашылығын, мазмұн межесін, яғни мағынасын өзгеріссіз екінші тілдің сөйлеу шығармашылығына айналдыру үдерісі», – деп атайды [150, 28]. Аударма түпнұсқаға сәйкес келу сипаты жағынан алтыға бөлінеді: *еркін аударма, эквивалентті аударма, дәлме-дәл аударма, аутентикалық аударма, канондық және расталған аударма*. Осының ішінде «Абай жолы» роман-эпопеясының орыс тіліндегі аудармаларын эквивалентті аудармаға жатқызамыз.

Осы уақытқа дейін «Абай жолы» романы орыс тіліне екі рет аударылды [151]. Оның алғашқысына автор М.Әуезов тікелей қатысқан. Сондықтан Л.Соболев бастаған аудармашылар тобы жасаған бұл аударма авторизацияланған аударма саналады. *Авторизацияланған аударма – шығыс мәтінді аудармашы түпнұсқаның авторымен бірлесе отырып аударған және автор оның даярлануын мақұлдаған аударма түрі* [150, 36]. Екінші рет романды А.Ким 2002 жылы орыс тіліне қайта аударды. «Абай жолы» романындағы бейвербалды амалдардың орыс тіліне баламалану ерекшеліктері автордың өзі аударуға ат салысқан алғашқы аударма нұсқасы негізінде талданады.

Кез келген көркем шығармадағы бейвербалды амалдар – әр халықтың психологиялық, танымдық, коммуникативтік, әлеуметтік ерекшеліктерінің тілде вербалдануының нәтижесі. Бұған дәлел ретінде «Абай жолы» романының орыс тіліндегі аудармасы мен түпнұсқасындағы тілдік емес құралдардың баламаларын атап өтуге болады. Мысалы, шығарма мәтіндегі «күлуді» білдіретін кинесикалық амалдар қазақ тілінде және орыс тілінде былайша вербалданады:

Күлу кинемасының орыс тіліндегі аудармасы

№	Түпнұсқа	Орыс тіліндегі баламасы	Аударма тілдегі параэлементтің мағынасы
1	-Абай сен де жүр!.. Атаң ғой, атаңа сәлем бер, - деп мысқыл еткендей сылқ-сылқ күлді. (I том, 98.)	-Абай, иди и ты тоже! Он же тебе тесть, отдай тестю сәлем! – и он иронически усмехнулся. (I т, 99 с.)	Усмехнуться. Слегка засмеяться (обычно с насмешкой, недоверчиво). [5]
2	Абай тағы бір рет, ай шеті қылдырықтай боп көріне түсіп, сөне қалғанда, еріксіз күліп жіберді (I том, 89 б.)	И когда он чуть заметной полоской опять сверкнула и мгновенно исчезла, он невольно усмехнулся (I т, 91 с.)	Усмехнуться. Слегка засмеяться (обычно с насмешкой, недоверчиво). [5]
3	Қараса, бұның шокпарын тартып ап, ақ боз аттың алдынан көлденең шығып, қазір шексілесі қатып, үнсіз күліп тұрған бағанағы шәкірт бала. (I том, 5 б.)	Не мерещится ли ему? Перед ним на коне – мальчик!.. Так вот кто налетел на него, отобрал шокпар, остановил коня и теперь заливается смехом, не в силах вымолвить слова. (I т, 32-33 с.)	Заливаться смехом. Смеяться громко, с перекатами голоса. [5]
4	Жанағы даурықтардың бәрін естіп, тоқыраған Шөженің өзі. Ол қатты, ашық дауыспен қарқылдап, сүйсіне күлді. (I том, 109 б.)	Услышав пререкания, Шөже остановился. Он громко рассмеялся, голос у него был звучный, ясный. (I т, 109 с.)	Громко смеяться. Начать сильно смеяться. Р. в глаза кому-н. (ответить смехом). [5] («Сүйсіне» сөзі аударылмай қалған. Бұл сөз эмоцияға жағымды реңк беріп тұр.)
5	Абай Майбасардың үп-үлкен басымен балаша қылжақтап отырғанына қызығып күле бастады. (I том, 115 б.)	Смелые шутки дяди понравились Абаю. Он рассмеялся. (I т, 115 с.)	Рассмеяться. Начать сильно смеяться. [5]
6	Бала топтан шыға бергенде, Айғыз күліп: - Пай, жаман қатындар сілекейлеп, баламыздың бетінен сүйер жер де қалдырмады-ау, - деп паңдана күлді де, Абайды көзінен сүйді. (I том, 11 б.)	– Ну вот, всякие грязнули заслюнявили все лицо нашему мальчику, и поцеловать некуда! – И она с высокомерной усмешкой поцеловала Абая в глаза (I т, 38 с.)	Усмешка. Улыбка, выражающая насмешку или недоверие, насмешливое движение рта. [5]
7	Үлкендер ақырын мырс-мырс күліп қойды. (I том, 15 б.)	Старики рассмеялись (I т, 42 с.)	Рассмеяться. Начать сильно смеяться. [5]
8	Дағдылы қорқақтығы, Құнанбайға қарсы шаба алмайтын ежелгі маймақтығы» деп, мырс етіп, мұрт астынан күліп қойды. (I том, 72 б.)	«Трусит по обыкновению, – подумал он. – Как всегда, слаб волей и боится Қунанбая...» И он едва заметно усмехнулся. (I т, 76 с.)	Усмехнуться. Слегка засмеяться (обычно с насмешкой, недоверчиво). [5]

9	Абай әлгі де қысылмай, күле қарап Майбасарды мысқыл еткендей мырс етті де, арт жағында тұрған домбыраны алып, даңғылдатып тарта бастады. (I том, 85 б.)	Но Абай и тут не смутился . Он ответил Майбасару насмешливой улыбкой , взял домбру, стоявшую за его спиной, и начал молча наигрывать. (I т, 116 с)	Улыбка . Мимика лица, губ или глаз, показывающая расположение к смеху или выражающая привет, удовольствие либо иронию, насмешку. [5]
10	Абай күлместен, мысқылдай қарайды . (I том, 13 б.)	Но Абай по-прежнему оставался невозмутимо серьезным, и только в глубине его глаз притаился добродушный смешок . (I т, 40 с.)	Смешок . Короткий, не сильный смех. [5]
11	Оспан оның тітіркенгеніне мәз болып , әкесін ұмытып сақылдап күліп : - Бақа! Бақа салып жібердім көйлегіне, - деп, Абайды бұрынғыдан да жаман тітіркендірді. (I том, 16 б.)	А Оспан, забыв о присутствии отца, покатываясь со смеху , закричал: – Лягушка! Я посадил ему за ворот лягушку! Абай забарахтался еще больше. (I т, 43 с.)	Покатываться со смеху . Громко, безудержно смеяться, хохотать. [5]

Қазақ көркем мәтіндерінде кейіпкер эмоциясын көрсетуде «мырс ету» параамалы көп қолданылады. Күлкінің бұл түрі қазақ танымына тән ерекшелік деуге де болады. *Жетінші, сегізінші, тоғызыншы* мысалдардағы «мырс ету» кинемасын аударма авторы үш контексте үш түрлі аударған. «*Дағдылы қорқақтығы, Құнанбайға қарсы шаба алмайтын ежелгі маймақтығы*» деп, *мырс етіп, мұрт астынан күліп қойды*» (I том, 72 бет) деген мысалдың аудармасында «мырс ету» кинемасы мүлдем аударылмаған. Ал аудармадағы *едва заметно усмехнулся* деген тіркес «мұрт астынан күлудің» баламасы болып тұр. Жетінші мысалдағы *мысқыл еткендей мырс етті* деген бөлік те аударма тілде *насмешливой улыбкой* болып біртұтас амал ретінде көрініс тапқан. Алайда баламада «мырс ету» амалына тән мағыналық ерекшелік айқын көрінбейді. Сегізінші мысалдағы *мырс-мырс күліп қойды* кинемалары аударма мәтінде ешқандай экспрессивті реңкін сақтамай, жай ғана күліп жіберуді білдіретін *рассмеялись* сөзімен баламаланған. Ал *мырс етуге* тән *кекету, болымсыз күлу, естілер-естілмес түрде күлу* деген секілді семантикаларды орыс тіліндегі *рассмеялись* сөзі бере алмайтыны байқалады.

Алтыншы мысалда «*паңдана күлді*» «*с высокомерной улыбкой*» болып аударылып тұр. Бұл аудармада параамал өзінің сипатын толық өзгерткен. Себебі «күлу» кинемасы кезінде беттің барлық бұлшық еттері қозғалып, эмоция дауыс арқылы да сыртқа шығады. Ал «улыбка» ол жай ғана езу арқылы орындалатын күлуге қарағанда қарапайым кинетикалық амал. Сонда контекстегі «паңдана» сөзі «с высокомерной» түрінде сөзбе-сөз аударылып тұрса, екінші компоненттің (күлу – улыбка) аударма мәтінде эмоциялық бояуы солғындаған. Керісінше төртінші мысалдағы *шек-сілесі қатып, үнсіз күліп тұрған* деген параамалды аудармашы өте жақсы аударған деп баға беруге болады. Орыс тіліндегі *заливается смехом, не в силах вымолвить слова* деген балама түпнұсқадағы эмоцияны толық көрсетіп тұр. Әрі мағыналық жағынан, әрі құрылымдық

жағынан аударма орыс тілді оқырман үшін түсінікті, қабылдауына жеңіл. Дәл осылай қазақ тіліндегі метафорланған бейвербалды амалдарды орыс тіліндегі мағынасы сәйкес келетін және эмоциялық әсері жоғары тұрақты метафоралы қолданыстармен баламалау аударма мәтіні үшін сәтті шешім болмақ. Аударма тілде көркемдік қуаты жоғары баламасын дәл тапқан балама ретінде соңғы мысалды айтуға болады: *сақылдап күліп – покатываясь со смеху*.

Шығарма мәтініндегі бір ғана «күлкі» кинемаларының аударылуына талдау жасау арқылы кейбір бейвербалды амалдарды аударудың қаншалықты күрделі, тіпті, барлық мағыналық реңкін аудармада беру мүмкін емес екені аңғарылды. Бұдан көркем мәтінге тән образдылық, көркем мәтіндегі тілдік бірліктерге тән ұлттық ерекшелік, күрделі семантикалық құрылымның аударма кезінде аудармашыдан біліктілік қана емес, жазушылық шеберлікті қажет етеді деген қорытындыға келуге болады.

«Абай жолы» романы – қазақ халқының жарты ғасырлық тарихын суреттейтін күрделі құрылымды роман. Сол себепті романның орыс тіліндегі аудармасында тілдік бірліктер (вербалданған бейвербалды амалдар) түрлі тәсілдер арқылы баламаланған. Түпнұсқа мәтіндегі ым-ишараттардың аударылу ерекшеліктеріне қарай бес түрге жіктеуге болады. Олар:

1. *Аударылмаған тілдік бірліктер (лақуналар);*
2. *Сөзбе-сөз аударылған тілдік бірліктер;*
3. *Сапалы аударылған тілдік бірліктер;*
4. *Теңестірілген тілдік бірліктер;*
5. *Еркін түрде аударылған тілдік бірліктер;*
6. *Жартылай аударылған тілдік бірліктер.*

1. Аударылмаған тілдік бірліктер (лақуналар) – шығарманың аударма мәтінінде түпнұсқадағы сөз, сөйлем, абзацтардың толықтай аударылмауына байланысты баламаланбаған параэлементтер. Мысалы, төмендегі сөйлемдердің орыс тіліндегі мәтінде аудармалары түсірілген (9-кесте):

Кесте 9

«Абай жолы» романындағы аударылмаған бірліктер

Түпнұсқа мәтін	Аударма мәтін
Бұл Бөжейдің <i>қабағын баққан</i> болатын. Өзі қалай көшерін айқын біле алмай, «соның бір білгені шығар» деген есеппен жалт берді. (I том, 21 б.)	-
Басының жалтыры да көрініп, Абай <i>еріксіз күлген-ді</i> . Асау тоқтамаған соң, шылбырды беліне орап алған Жұмабай, шалқайып табандап тартып, көше бойының құмын боратып, тағы да рәсуа болған. Бұл тұста Сағитқа қосылып Абайдың да <i>шексілесі қатып еді</i> . Жұмабай көк құнанды зорға тоқтатып алған соң, Абай Сағитты маңайлатпай қуып жіберіп, сол бәледен Жұмабайды да, құнанды да өзі құтқарып шыққан. (I том, 24 б.)	-

2. Сөзбе-сөз аударылған тілдік бірліктер – түпнұсқа мәтіндегі сөздердің аударма мәтінде мағынасы мен формасы толық сәйкес келетін абсолютті эквиваленті арқылы баламаланған параэлементтер. (10-кесте)

**«Абай жолы» романындағы сөзбе-сөз аударылған
тілдік бірліктер**

<p>- Жан ашырдың жоқ болғаны ғой! - деп, бағанағы көпшілік тағы қозданайын деп еді, Құнанбай Сүйіндікке қадалып тұрып, қамшысын жаңағы сөйлегендер жаққа нұсқап: – Тоқтап, аналарыңды! - деп, ақырып қалды. (I том, 67 б.)</p>	<p>Кунанбай быстро смирил их. Его единственный глаз впился в Суюндика, и, указывая плетью на шумевших, он властно крикнул: – Уйми их! (I т, 71 с.)</p>
<p>Қаны қашып, аппақ сұр болған бетін төр жаққа, атасы жаққа бұрып, қадала тыңдады. Екі көзі жасқа толып, қызарып тұр. Аш бетінде көк тамырлары білінеді. (I том, 27 б.)</p>	<p>Бледная, повернув бескровное лицо к свекру, она чутко прислушивается. Покрасневшие глаза полны слез. На висках похудевшего лица чуть заметно бьются голубоватые жилки. (I т, 54 с.)</p>
<p>Үндемей біраз тұрып, қабағын шытып, қатты күрсінді де, айналып жүріп кетті. (I том, 26 б.)</p>	<p>Несколько минут мальчик стоял молча, нахмуриев брови, потом круто повернулся и пошел, тяжело вздохнув. (I т, 53 с.)</p>

3. Сапалы аударылған тілдік бірліктер – түпнұсқа мәтіндегі параэлементтің семантикалық, экспрессивтік реңктері толық сақталған және аударма тілде кең тараған қолданыстар арқылы баламаланған түрлері. (11-кесте)

**«Абай жолы» романындағы сапалы аударылмаған
бірліктер**

<p>Қараса, бұның шокпарын тартып ап, ак боз аттың алдынан көлденең шығып, қазір шексілесі қатып, үнсіз күліп тұрған бағанағы шәкірт бала. (I том, 5 б.)</p>	<p>Не мерещится ли ему? Перед ним на коне – мальчик!.. Так вот кто налетел на него, отобрал шокпар, остановил коня и теперь заливается смехом, не в силах вымолвить слова. (I т, 32-33 с.)</p>
<p>Қодар қиналып, қатуланған соң Жәмпейіс өз естігенін тегіс айтқан жоқ. (I том, 29 б.)</p>	<p>Увидев потемневшее лицо Кодара, Жампеис понял, что лучше не растревлять его раны и не рассказывать дальше. (I т, 56 с.)</p>

4. Теңестірілген тілдік бірліктер – аударма кезінде грамматикалық трансформация аудармадағы қосып жеткізуді (синтаксистік және лексикалық себептер бойынша) енгізетін әрі қалдырып кету (аудармашы назарынан және аударма тілінен тыс қалған аударма ақпаратын шығарып тастау), қосып жеткізу үйлесімділігі нәтижесінде баламаланған сөздер. (12-кесте)

«Абай жолы» романындағы теңестірілген тілдік бірліктер

Осы кезде Жиренше Абайға ым қақты да, тысқа шығып кетті. (I том, 23 б.)	Воспользовавшись этим Жиренше подмигнул Абаю и кивнул на дверь юрты. (I т, 51 с.)
Осының бәрін көп күрсініп отырып айтқан Ұлжан Абайға үлкен ой салды. (I том, 246 б.)	Ұлжан говорила об этом вздыхая. Абай мрачно насупил брови и глубоко задумался (I т, 414 с.)

5. Еркін түрде аударылған тілдік бірліктер – түпнұсқа мәтіндегі параэлементтің аударма мәтінде аудармашы тарапынан әрі мағыналық, әрі формалық жағынан өзгертіліп, кеңейтіліп, түсінідіріле отырып баламаланған тілдік бірліктер. (13-кесте)

«Абай жолы» романындағы еркін түрде аударылған тілдік бірліктер

Әрқашан жан күйер жақын қабырына ыстық-ыстық жастар төгеді. Қайта-қайта бас қояды. Үндемей иіндері тийісін, көздерін алмай, ұзақ-ұзақ отырысады. (I том, 31 б.)	Приходя на могилу, каждый из них всегда мысленно творил свою собственную молитву: делился своим горем с Кутжаном и тихо упрекал его, зачем он их покинул, и слезы их скатывались на могилу. По многу раз они припадали к ней лицом и, прижавшись друг к другу, молча сидели, не отрывая от нее взгляда. (I т, 58 с.)
Сондықтан, бар тапқаны - ол бала өзі еместей, үндемей сазарып қалды. (I том, 15 б.)	Абай старался сделать вид, будто все, о чем здесь говорят, не имеет к нему никакого отношения. (I т, 42 с.)
Құнанбай кеп түскенше, жапырлап сөйлеп, күлісіп, әзілдесіп жатқан әйелдер, Құнанбайға Майбасар кеп есік ашқанда, жым-жырт бола қалды. (I том, 59 б.)	Как только в двери, почтительно откинутой Майбасаром, появилась внушительная фигура Кунанбая, смех и шумная болтовня женщин мгновенно смолкли. (I т, 85 с.)
Үндемей түйіліп, сұп-сұр боп отырған Құнанбай ол сөздерге көп шіміркенген жоқ. Қайта, теріс көргендей мойын бұрып, Абайға қарап: – Одан да, не күтсеңдер де осы жаман қарадан күтсеңдерші! – деді. (I том, 14 б.)	Но он сидел молча, не выражая никакого удовлетворения от расточаемой лести. Вдруг, как бы наперекор всем, он проговорил, повернувшись в сторону Абая: – Если уж чего-нибудь ждать – так ждите только от этого черномазого мальчугана! (I т, 42 с.)
– Абай қайда өзі? – деп еді, Ұлжан Абайдың жатып қалғанын айтып, Жиреншеге біраз наразы салқын жүзбен қарады. (I том, 44 б.)	Под конец он спросил: – А где же он сам? Ұлжан ответила, что Абай спит, и холодно посмотрела на юношу. В голосе ее звучала досада. (I т, 70 с.)

6. *Жартылай аударылған тілдік бірліктер* – параэлементтердің аударма мәтінде белгілі бір бөлшегі аударылмауы немесе түсірілуі арқылы баламаланған түрлері. (14-кесте)

Кесте 14

«Абай жолы» романындағы жартылай аударылған тілдік бірліктер

<p><i>Еркектер төрге кеп отырған соң</i>, тек Таңшолпан ғана: – Мына шешелерің, жеңгелерің сыбаға әкеп, енді тарағалы отыр, қарағым. Кәрі әжеңнің орны бір басқа. Оныкі өз бетімен барып жатыр! - деп, Зерені де есіне алды. (1 том, 59 б.)</p>	<p><i>Мужчины сели</i>. Обратиться к Кунанбаю осмелилась одна Таншолпан, вторая жена его отца, считавшаяся старшей после Зере: – Сын мой дорогой, твои матери и старшие невестки принесли угощение. Теперь мы собираемся расходиться. А твоей старой матери приношение уже сделано... (I т, 85 с.)</p>
<p>2) - Бүйтпеуші еді. Бұл неғып жүр? Әлде, күзеуінің оты тозды ма екен? – дегенде, Бөжей бұған <i>барлай қарады</i> да, мырс етті. (1 том, 57 б.)</p>	<p>Никогда раньше Кунанбай так не делал... Кормов ему не хватило, что ли? Божей <i>взглянул</i> на него и рассмеялся. (I т, 60 с.)</p>

«Абай жолы» романындағы бейвербалды амалдар кинесикалық, просодикалық, проксемикалық, такесикалық топтарға ажыратыла келіп, этностың салт-санасын, әдет-ғұрпын, ой-танымын бейнелейді. Олар көркем шығарма тілінде, адами қарым-қатынаста айтайын деген ойды жан-жақты, терең жеткізуге ықпал етеді. Ал бейвербалды амалдардың аудармасын зерттеу нысанына алу шығарманың басқа мәдениет өкілдеріне қандай сипатта жетіп жатқандығын және қазақ тіліндегі бейвербалды амалдардың вербалдану ерекшеліктерін көрсетуге мүмкіндік береді. Сол арқылы аударма тәжірибесінде параэлементтерді қандай тәсілдер арқылы баламалаған тиімді екенін де салыстырмалы түрде анықтауға болады. Роман аудармасында баламаның төмендегідей түрлері мол ұшырасады:

- Толық балама
- Абсолютті балама
- Контекстуалды балама
- Формалды балама
- Баламасыз лексика
- Ішінара сәйкестік
- Лексикалық алмастырулар

Көркем мәтіндердің аудармасы – әлемді мәдени жағынан танудың, адамзат баласының ұжымдық зейінін кеңейтудің құралы, сондықтан көркем аударманы мәдениетаралық коммуникацияның ерекше – мәдени-этникалық және көркем-эстетикалық түрі болып табылады. Сол себепті аударма сапасы мен көркем шығарма құндылығын сақтауда әр ұлттың танымдық ерекшеліктері, тілдік бірліктердің мағыналық және құрылымдық жақтан күрделілігі, аудармашы құзыреттілігі, аударма тәсілдері басты назарға алынуы керек.

5.4 Параллель ішкорпусқа алынған аударма мәтіндерді талдау

Параллель корпус – аударма мәтіндерді теңестіру арқылы жинақталатын база. Параллель мәтіндерді әзірлеуде *Excel* форматында қолма-қол мәтін теңестіру тәсілімен қазақ тіліндегі орысшаға аударылған шығармалардың не сөйлемінің, не тұтас текшесінің (абзацының) тұсына орыс тіліндегі аудармасы табылып, орналастырылады. Бұл жұмыс аударма сапасының да назардан тыс қалмай, мұқият зерделенуіне мүмкіндік берді. Кеңестік кезеңде жасалған аудармалар мен қазіргі тәуелсіз Қазақстан тұсында жасалған аудармалардың айтарлықтай айырмашылығы бар екені көрінеді. Әсіресе бұрынғы көркем аудармаға гегемонияның (саясат үстемдігінің) тигізген залалынан бүгінгі күнгі аудармаларда толықтай айыққанымыз көрінеді.

Аударма өндірісінің сапаға бет бұрып, қарқынды дамуы әлемдік деңгейде де көрініс тауып келеді, оның бір айғағы – уақыт өткен сайын жетілдіріліп келе жатқан Аударма стандарттары. Бұған ISO 9000 сериясының стандарттары кіреді, олар сапа менеджменті қағидасымен жұмыс істеу кезінде тұтынушылар мен мүдделі тараптардың қажеттіліктерін қанағаттандыруға арналған. ISO 9001 стандартқа сай болғысы келетін ұйымдар орындауы тиіс талаптарды қарастырады және миллионнан астам ұйымды сертификаттап отырған бүгінгі таңда кеңінен қолданылатын басқару құралдарының бірінен саналады.

Басында өндірістік кәсіпорындар үшін әзірленген аударма сапасының стандарттары біртіндеп бірінің орнын бірі басып, сапа менеджментіне қызығушылық артқан сайын ауық-ауық жаңартылып отыратын, реттеуші құжаттарға айналды. Олардың қатарында немістің DIN 2345, Италияның UNI 10574, Австрияның Önorm D 1200 мен Önorm D 1201 және Канаданың CAN CGSB 131.10 ұлттық стандарттарын атауға болады.

Өткен XX ғасырдың 90-жылдары сапа стандарттары мен кәсібилік ұғымдары адамдардың психикасында болмаған. Соның салдарынан тапсырыс берушілер мен қызмет көрсетушілер арасында түсінбеушілік жағдаяттар орын алып, екі тараптың қақтығысқа дейін баруына жол берілген. Бұл келеңсіздіктер аудармада сапа стандарттарын әзірлеудің қажеттілігін тудырды. Неміс халқы бірінші болып DIN 2345 сапа стандарттарын әзірлеп, бекітті. Құжат 1998 жылы өздерінің автомобиль өнеркәсібіне арналған нұсқаулықтарын жасау үшін өндірістік және техникалық талаптарын қанағаттандыру мақсатында әзірленген. Бұл тәсіл өзінің оң нәтижесін көрсеткен соң әлемге тез арада таралып, ұлттық стандарттардың пайда болуына, аударма қызметіндегі сапа стандарттары бойынша нұсқаулықтардың дайындалып, күшіне енгізілуіне себін тигізді.

Еуропалық EN 15038 аударма қызметі стандарты 2006 жылдың 1 тамызында күшіне енді де, бұрынғы CEN (the European Committee for Standardization) 30 мүше-елінің стандартының күшін жойды. Ол аударма саласында қолданылатын терминологияны бір жүйеге келтіруге, тіл қызметін көрсетушілерге қойылатын негізгі (адами, техникалық ресурстар, сапаны бақылау деген сияқты т.б.) талаптарды анықтауға және тұтынушылар мен қызмет көрсетушілердің міндеттері мен құқықтары тұрғысынан өзара қарым-қатынасқа бағытталған. EN 15038

стандарты да 2015 жылы одан әрі қарай жетілдіріле түскен ISO 17100:2015 стандартына жол берді. Айтарлықтай айырмашылық болмаса да, кейінгі құжат аудармашының біліктілігі мен жазбаларды сақтау мәселесі тұрғысынан толықтырылған (<https://www.modlingua.com/blogs/757-quality-standard-in-translation>).

Бүгінгі күнгі аудармашы мамандар үшін, олардың аударма бизнесі үшін осы стандарттарды білу маңызды. Құқықтық, медициналық аудармалар сияқты аударманың нақты салаларымен айналысқанда, неғұрлым нақты терминологияны білетін коммерциялық немесе техникалық аудармалар жасалады және бұл салалар стандартталған құжаттармен жұмыс істейді. Аударма стандарттарының қазіргі индустриялық әлемде қалай жұмыс істейтінін білгеннен кейін бұл стандарттар туралы хабардар болмаудың өзі аудармашылар мен тіл мамандарының кәсіби дамуына қалай кедергі жасайтынын оңай түсінуге болады. Әрдайым ізденіс үстінде жүретін, осы стандарттар бойынша білімін жаңартып отыратын сертифицикатталған ұйымдар мен мамандар басқаларға қарағанда әлдеқайда сенімді, әрі сапалы жұмыс істейтіні анық.

Егер аудармашы ізденімпаз, жаңашыл болса, халықаралық деңгейдегі осындай талаптардан хабардар болады. Ал мұндай стандарттардың бар болуының өзі аудармашыны дұрыс жолға бағыттап, қате басуына жол бермейді. Одан өзге бірқатар ғалымдардың еңбектерінде аударма сапасын ашатын өлшемшарттар төмендегіше болып келеді:

- Түпнұсқаның мағынасы ашылды ма?
- Емін-еркін оқыла ма?
- Грамматикалық немесе орфографиялық қателер бар ма?
- Консистенция, яғни ішкі бірізділік, үйлесімділік сақталған ба?

Осындай сұрақтарға жауап іздей келе, зерттеулер «жоғары сапалы аударма түпнұсқа мәтінінің мағынасын ашқан, дұрыс жеткізілген, қатесіз және ішкі үйлесімділігі бар аударма» деген қорытындыға тоқтайды (D. Brown “Pacific International Translations” <https://www.pactranz.com/translation-quality>).

Олай болса, осы орайда бүгінгі көркем аударма ішкорпусын жасау барысында дереккөз ретінде нысанаға алынған шығармалардың аудармаларына талдау жасадық.

Аударма теориясында аударылуға тиісті материал аударылатын және аударылмайтын транслатема болып екіге бөлінеді. Еске сала кетейік, транслатема – аударма үдерісіндегі тілдік бірлік. Бұлардың ішінде аудармашыны әрдайым тығырыққа тірейтін транслатема тілден тілге аударылмайтын этнографизмдер – этномәдени тілдік бірліктер болып келеді, олар реалий атаулар ретінде де белгілі. Оларды тілден тілге көшірудің арнайы амал-тәсілдерін меңгеруіміз қажет болады.

Аударма ішкорпусын жасау барысында түпнұсқа мәтініне аударма мәтінін қолма-қол теңестіру кезінде біздің назарымызға іліккен екі жағдаят болды. Оның бірі – қазақтың қара домбырасының аудармада берілу жайы болса, екіншісі – аударма мәтінінде ойдан шығарылған ақпараттың пайда болуы.

Қазақтың қара домбырасы С. Елубайдың «Ақбоз үй» трилогиясының «Мінажат» романында бірнеше жерде кездеседі екен, соның бірі (контексімен алайық):

«...Айтпақшы, әлгі күйші қайда? Тебіренген жанын тербетіп әкететін бір ырғақ, бір күй керек-ақ. Сенделіп жан-жағына қарады. Керегеге сүйеулі тұрған қара домбыраға ұмтылды. Қолға алды. Қарағай сапты күмбірлеген қара домбыра. *Құлағын бұрады. Бұрауын түсірді. Қазір* қара домбыра сөйлей жөнейтіндей, шекке тиер-тиместен боздай жөнейтіндей еді, олай болмады. Домбыра «дың» еткені болмаса, сөйлемеді».

Аудармасы (аудармашылар: Л. Космухамедова және А. Жақсылықов):

«...Кстати, где же обещанная домбристика? *Сейчас ему ой как нужен кюй, музыка, чтобы закачала она душу волнами ритма. Встал, маетно заходил туда-сюда. Увидел темную домбру, висевшую на кереге юрты. Взял ее в руки. Это был сработанный из сосны звучный инструмент. Подкрутил ушки. Сейчас бы ударить по тугим струнам, уйти в забытье в рокоте музыки, однако душевный порыв тут же погас.* Домбра издала только жалобный стон».

Көретініміз, осы контексте түпнұсқада «домбыра» сөзі 1 рет, «қара домбыра» тіркесі қатарынан 3 рет кездеседі. С.Елубай танымал, шебер жазушы бұлай қолданса, қара домбыраға салып отырған өзіндік «салмағының» болғаны. Аудармада «домбра» 1 рет, ал қара домбыра «темная домбра» болып 1 рет беріледі. Бұл жердегі назарымыз, қара домбыраның неше рет берілгенінен де бұрын, орыс тіліндегі мәтінде «темная домбра» болып берілгеніне көбірек ауды.

Этнографизмнің келесі кездескен жері:

«Қолындағы қарағай сапты қара домбыра – Хансұлудың өз домбырасы. *Бүгін Ждақай алдыртқан Сұржекей келеді деп. Бұны да алдыртып отырған Ждақай, «Егер Шегенің жайын білгің келсе, осыдан оңтайлысы жоқ» деп. Бұл елге аты әзірейілдей үрейлі Сұржекей дастарқан басында күй тыңдағанды қалайды-мыс. Хансұлу бұларға сол үшін керек болыпты-мыс».*

Аудармасы:

«В ее руках небольшая темная домбра, ее собственная домбра. Накануне Ждахай забрал ее, сославшись на приезд Суржекея. Он сказал: «Если ты хочешь узнать что-нибудь о Шеге, нет более удобного случая». И она поняла, что небезызвестный Суржекей, которого аульчане боялись, словно Азраила – ангела смерти, желает насладиться музыкой за дастарханом».

Мұнда да «қара домбыра» – «темная домбра». Ал жалпы мәнмәтіннің берілу сарынына қарап, аудармашыны осал деуге келмейді.

Келесі қолданыс мынадай контексте беріледі:

«...Құлақ күйі келген кең құрсақ қара домбыра енді тақымға тиген жарау аттай зымыран шабыс тілеп лып-лып етеді. Ақтарыла жөнелгісі кеп күмбірлеп құлшынады. Көкірегінде көптен шер болып шеменделіп жатқан ащы запыран домбыраның тілімен ағындап тұма бұлақша атылғысы бар. Неге екені белгісіз, домбыра – жансерігі сыңғыр қара домбыра «Қайран шешем» деп, боташа теңселе жөнеді. Көкіректі тепкілеген зәбір-жапа күйге бөленіп, атылып сыртқа шыққанына жаны жусай бастады. Мына қанын ішіне тартып, сұрланған, түсі суық жыландай жанмен тілдесуден тайсақтағаны, қашқаны еді... келе домбыраға жармасқаны. Күйге кезек бергені...»

Аудармасы:

«Настроенная для кюя, широкая и круглая в нижней части, черная домбра начала подрагивать в ее руках, словно застоявшийся скакун, просящийся в

стремительный намет. И тербит, и рвется разлиться потоком, гудит горловым, гортанным голосом. И давняя тоска, черной желчью палящая нутро, бьется теперь, чтобы вырваться кюем, клокочет, чтобы ударить родником, нашедшим ход в каменной теснине. Но кто знает, отчего голос черной домбры, *подруги ее души, на сей раз вырвался плачем верблюжонка, ищущего потерянную мать. И обида, сжимавшая грудь, пробила путь и вырвалась глухим стоном струн, гулким кюем, в котором ее мятущаяся душа теперь стремилась сбросить оковы. Она, затаенная во хмари, и хоронилась бы там, будто напуганная змеиным холодом суровой души, но... домбра позвала... кликнула властным голосом кюя».*

Бұл жердегі түпнұсқа мәтiнiнде «домбыра» – 5 рет, «кара домбыра» – 2 рет кездеседi. Аудармада «домбра» – 1 рет, «кара домбыра» ендi «черная домбра» болып, 2 рет қолданылады.

Тағы бiр мысал:

«...Дауылды күн ыңыранып, арнасына сыймай тебiренген теңiздеi кең тынысты бiр көкiрегi қаяулы ырғақ... карағай сапты кара домбыраның кәрi көкiрегiнен суылдап бас көтердi. «Кiшкентай» едi бұ күй. Шамырқанған шатақ күй. Күй көп ұзамай, карашаның ауыр кара бұлтындай түйiлдi. Кең дүниенiң ысқырып, апшысын қуырды...».

Аудармасы:

«... Сизо-черная буранная ночь с рыданием и стоном хлещет по каньону сурового времени, и печалится и бьется в сердце объемный ритм безмерного плача... с пепельным шелестом и шурушанием поднимает голову древний напев, разбуженный силой черной домбры. Это был кюй «Кiшкентай». Кюй яростный, неукротимый. Спустя некоторое время, музыка сгустилась на одной ноте, мрачной осенней тучей зависла в воздухе. Вдруг, оглушая мир свистом, она понеслась, завихрилась.»

Мұнда да «кара домбыра» «черная домбра» болып аударылғанын көрiп отырмыз.

Шығармадағы келесi қолданыс:

Хансұлу өңi қуқылданып, кiшкентай танауы қусырылды. Түндеi тұнық кара көзi оттай шоқтанды. Саусағы сүйкенiп едi... күмбiр-күмбiр еттi домбыра. Кара домбыра көп ұзамай, құдiреттенiп күңiрене бастады.

Аудармасы:

На щеках Хансулу вспыхнул румянец, возбужденно расширились ноздри. Бездонные черные глаза заискрились смоляным огнем. Едва пальцы прикоснулись к струнам, как гулко отозвалась домбра. Черная домбра не заставила себя ждать, она затрепетала, запела, зарыдала.

Романдағы кара домбыраның соңғы кездесетiн жерi:

«Қарағай сапты кара домбыра «Қайран шешем!», «Қайран шешемi» деп боздап, кең дала боп көсiлiп, самал жел боп есiлiп, отырғандарды мына қапас кара түн, *тар* қапастан алып шығып, аруана сұлу дүниенi аралатып әкеттi».

Аудармасы:

«И мелодия черной домбры, *заливаясь стенаниями дитя по потерянной матери, обдавая просторы рыданиями, затрепетала порывистым ветром, она распахивала перед слушателями другой мир, поднимая их души из мрачных и душевных темниц в дивный мир могучей красоты».*

Сұрақ: «Қара домбыраның» бірде «темная домбра», бірде «черная домбра» болып берілуінен лексеманың мәдени семантикасы ашылды ма?

Семасиологиядағы көпмағыналық (полисемия) ұғымының бір мысалы – осы қазақ тіліндегі қара сөзі деуге болады. Қара жер, қара шаңырақ, қара орын, қара орман, қара халық, қара қазан, қара домалақтар, қара шал, қара кемпір, қара домбыра, т.б. дегендегі қара сөзі жәй, қарапайым деген мағынаны бергенімен, оның түпкі мәдени танымдық негізінде қастерлі, қасиетті, киелі деген семантика жатады. Оны дәлелдеу қиын емес, осы тіркестердің кез келгенімен тілімізде жиі кездесетін мысалдар алайық:

Қара жерді дәретсіз баспаған әулие адам.

Қара шаңырақтың отын сөндірмей отырған әулеттің кенжетайы осы.

Әкем артында қалып бара жатқан қара орманын қимай, ұзақ жатты.

Қара қазан оттан түспесін.

Қара домалақтар көбейе берсін.

Қара халықтың ортасынан шыққан депутат.

Қазақ үшін қара домбыраның үніне жететін не бар дейсің?!

Қара шаңырақта қара шал мен қара кемпір қауқиып отырса, не арман бар?

Осындай күнделікті қолданыстағы коммуникациямыздан қара сөзінің қастерлі, қасиетті, киелі деген реңкін анық көруге болады. Сондықтан, қара домбыраны орыс тіліндегі мәтіндерде «темная домбра», «черная домбра» деп калькалау тәсілімен аудару жөн болмайды. Қандай болғанда да, бұл жерде домбыраның түсі айтылып тұрған жоқ, оған қоса қазақ домбырасының түсінің қара болуы да үйреншікті жай емес. Сондықтан, мағынаны толық жеткізу үшін қара домбыраны «сокровенная домбра» деп аудару керек деген ой қорытамыз, сонда орыс тілді оқырманға оның қастерлі деген мағынасы жеткізіледі. Қазақтың қасиетті қара домбырасы деген қолданыс та бекерден-бекер айтылмаса керек.

Айта кету керек, қара сөзінің қара қалам, қара шаш дегендегі денотат мағынасы мен ақ-қарасын ашу, қарасын батыру, қара аспанды төндіру, қара түнек, қара ниет т.б. толып жатқан коннотат мағыналарын ашу бұл мақаланың міндеті болып саналмайды.

Мәтін теңестіру барысында назарымызды аударған екінші мәселе – Ш.Бейсенованың «Мизамшуақ» әңгімесі мен бұл шығарманың орыс тіліндегі А.Жақсылықов аударған «Бабые лето» деп аталатын нұсқасы.

– Несін жасырайын, бағана сізді «Дастарханның» алдында алғаш көргенде бір құрдас жігіттің келіншегіне қатты ұқсатып, сол екен деп бұрылып барғаным рас.

Бұл – шығарма желісінде көлік жүргізіп отырған профессордың бір бейтаныс әйелді үйіне әкеліп тастап тұрғандағы сөзі. Орыс тіліндегі аудармасы түпнұсқасына қарағанда ұзақтау:

– Не буду скрывать, возле «Дастархана» я издали принял вас за жену одного своего друга, поэтому завернул к вам. Подбросив её до дома, хотел поддеть своего друга, дескать, почему заставил ходить жену пешком? Тем самым хотел раскритиковать его на поллитровку.

Расында, сөйлемнің көлемі аудармада екі есе ұлғайған, екінші жартысы ойдан қосылған. Одан әрі қарай, «Е-е-е, бәсе» деп ойлады Ақжамал» деген бір

ауыз сөзді аудармай, оның орнына аудармашының жалғаған диалогы тіпті таң қаларлық:

– *Жаль, потеряли вы бутылку!*

– *Ничего страшного...*

Он внезапно повернулся к ней, в его глазах заискрилась веселье:

– *Может быть, вдвоем разопьём?..*

– *Я с незнакомыми людьми вино не пью.*

– *Давайте тогда познакомимся.*

– *У меня нет привычки знакомится с людьми в транспорте.*

– *Ну-у-у, что это за свод законов. Вы ходите только по линии законов? А за черту туда или сюда выйти нельзя?*

– *Нельзя!..*

«Ну, вот ты и открылся» – подумала Акжамал.

Енді сұрақ туындайды. Мұндай ақпаратты қосу не үшін қажет? Ішімдік атаулы сөз болмаған мәтіннің арасына адамға да, Жаратушыға да жақсы көрінбеген амалды ойдан жасап қосу және оны қоғамның ең жоғарғы әлеуметтік қабатының өкілі, профессорға таңу не үшін қажет болды екен деген сұрақ тұрады.

Жоғарыда аталған халықаралық, ұлттық стандарттардың бірінде де аударма мәтініне ойдан құралған ақпарат, қосалқы жағдаят немесе оқиға қосу дегеннің бірі де қолдау таппайды, себебі бұл ақылға сыймайтын нәрсе. Ал аса қажеттілік жағдайында, мейлінше қысқа қосалқы мағлұматтың керек болатыны түсінікті. Мысалы, реалия атаулардың түсіндірмесі берілмесе болмайды, бұл – қажеттілік.

Кеңестік кезеңде жасалған І.Есенберлиннің «Көшпенділер» трилогиясының орыс тіліндегі, орыс тілінен жасалған ағылшын тіліндегі аудармаларын зерттеу барысында мұндай мысалдарды молынан кезіктірген едік. Бірақ ол кезде бүгінгідей аудармашыларға қойылатын талаптар (бар болғанмен) ақпараты қолжетімді бола бермеген, бір мәдениетке екінші мәдениеттің үстемдік көрсету деректері де көптеп кездесіп жататын уақыт болатын. Оның үстіне ол кездегі аудармалардың көбі өзге ұлт өкілдері арқылы жасалатын. Ал бүгінгі біз кезіктіріп отырған жағдайдың тағы бір түсінбеушілік тудырып отырған тұсы – аудармашының қазақ ұлтының өкілі және тәжірибелі аудармашы екендігі.

Қорыта келгенде, мәтін теңестіру арқылы алынған материал, яғни түпнұсқа мен аударма мәтіні қазақ тілінің корпусын жасауға «қаланып жатқан құрылыс материалы» болғандықтан, оларды жарамды мен жарамсыздық тұрғысынан сұрыптаудың да қажеттілігі сезіледі.

5.5 Параллель мәтіндер корпусының лексикографиялық әлеуеті

Дж.Веронис «Розетта тасынан ақпараттық қоғамға дейін: параллель мәтіндерді өңдеу әдістеріне шолу» [152] атты мақаласында мәтіндердің параллельді корпусын екі немесе одан да көп тілге аударылған құжаттар жинағы ретінде қарастырады, сонымен қатар ол параллель мәтіндер корпусының метамәтіндік, мәтінаралық және морфологиялық белгіленімнің болуы, сондай-ақ мұндай корпусқа репрезентативті фрагменттердің емес, толық мәтіндердің

қосылуы сияқты ерекшелігін атап өтті. Мәтіндердің мұндай жиынтығы аударма сөздіктер жасау саласындағы лексикографиялық жұмыстарды автоматтандыру үшін, сондай-ақ лексикология және фразеология саласындағы контрастивті зерттеулер үшін пайдаланылуы мүмкін. Қостілді және көптілді лексикографияда параллельді мәтіндерді қолдану:

– бастапқы мәтіндерде қолданылатын сөз тіркестерінің мағыналарын талдау арқылы сөздікке енген аудармалар жинағын байыту;

– терминологиялық сөздіктерге жиілікті және тұрақты конструкцияларды енгізу үшін белгілі бір пәндік саладағы мәтіндердегі нақты сөздер мен сөз тіркестерінің қолданысы мен мағыналарын нақтылау;

– екі тілді және көптілді сөздіктерде бұрыннан тіркелген лексикалық бірліктердің мағыналарын, әсіресе идиомалар мен терминологиялық тіркестерге қатысты тексеру;

– машиналық аударма жүйелерінің автоматты сөздіктеріне және глоссарийлерге енгізуге ұсынылатын тұрақты сөз тіркестерін және идиоматикалық тіркестерді ажыратуға мүмкіндік береді.

Осылайша, параллель мәтіндердің толық мәтінді деректер қорын талдау негізінде «бастапқы лексикалық бірлік – аударма» сияқты сөздердің тұрақты жұптарын анықтауға болады, сонымен қатар мұндай таңдауды автоматтандыруға болады.

Әрине, әртүрлі мәтіндерден лексикалық бірліктердің жұптарын айыру үшін, лексикографиялық мақсаттарда параллель корпусты пайдалану үшін оны алдын ала теңестіру қажет.

Мәтіндерді сөйлемдерге теңестіру күрделі жұмыс болып табылады, оның бірнеше шешімі бар:

– мәтінді сөйлемдерге сегменттеу мәселесін шешудің күрделілігі (тыныс белгісі ретіндегі нүктенің көп мағыналылығы, көркем әдебиет пен публицистика мәтіндеріндегі тікелей сөйлеуді сипаттау ерекшеліктері, тақырып болған жағдайда сөйлемнің соңын бекітудің болмауы және т. б.);

– мәтіндерді енгізгенде басы мен аяғының сөйлемдерге бөлу кезінде сәйкес келмеуі, мұндай сәйкессіздік мәтіндерді қолмен аударғанда пайда болады және сөйлем шекараларының сәйкес келмеуінің келесі алты нұсқасында көрінеді.

Мәтіндерді теңестіру аударма модельдері арасында алты ықтимал сәйкестігі бар болжам негізінде жүзеге асырылады:

– бір сөйлем бір сөйлеммен аударылады;

– екі сөйлем бір сөйлеммен аударылады;

– бір сөйлем екі немесе бірнеше сөйлеммен аударылады;

– екі сөйлем екі сөйлеммен аударылады, бірақ олардың ішкі шекаралары сәйкес келмейді;

– түпнұсқа мәтіннің сөйлемі аударылмайды;

– аударма мәтніндегі сөйлемнің түпнұсқасында баламасы жоқ және оны аудармашы енгізеді.

Теңестіру процедурасын автоматтандыру кезінде мәтін абзацтарының сәйкестігіне сүйене отырып, осы модельдерге сәйкес келетін жұптар бөлінеді. Демек, параллель мәтіндерді сөйлемдер бойынша автоматты түрде теңестіру

әрдайым қолмен өңдеуді қажет етеді, оның шеңберінде сөздердің, сөз тіркестерінің, синтаксистік конструкциялардың ықтимал сәйкестіктерін іздеудің шекаралары белгіленеді.

Лексикографияда параллель мәтіндерді қолданудың кең тараған нұсқаларының бірі білім беру мақсатында да, лексикографиялық зерттеу үшін де конкорданстарды құру болып табылады. Мысалы, Бирмингемде әзірленіп жатқан көптілді параллель конкорданс халықаралық жоба ретінде құрылуда, жобаға Еуропаның 6 елінен 6 университет қатысады. «Коллинз Бирмингем университетінің халықаралық тіл дерекқоры» [153] деп аталатын жоба көркем әдебиет және техникалық мәтіндер корпусынан дат, ағылшын, француз, неміс, грек және итальян тілдеріне параллель конкорданстар жасайды.

Бұл конкордансты жасаудағы мақсат – тіл үйрету болғандықтан, әр тілде классикалық көркем шығарма таңдалып, аудармалары теңестіріледі. Аудармаларды таңдауды және олардың сәйкестігін және/немесе баламалылығын бағалауды жоба авторларының өздері жүзеге асырады.

Егер мұндай корпусты екі немесе одан да көп тілге аударылған құжаттардың жиынтығы деп анықтайтын болсақ, аудармаларды таңдау принциптерімен байланысты жұмыстар жасалса, параллельді мәтіндер корпусының сапасы мен оның лексикографияға жарамдылығы арта түседі.

Бұл анықтамада мәтіндердің параллельділігін анықтайтын аударма операциясы қызығушылық тудырады. Себебі, лексикографиялық ақпаратты алу үшін параллель мәтіндер корпусын пайдалану тек сөйлемнен төмен деңгейде мәтінді теңестірудің дұрыстығымен ғана емес, сонымен қатар теңестірілген мәтіндердің эквиваленттілігімен де анықталады.

Осы тұрғыдан алғанда, мәтіндердің параллельді корпусындағы түпнұсқа мен аударма арасындағы әртүрлі қатынастарды анықтауға болады. Оларды ретімен қарастырайық.

Түпнұсқаға түпнұсқалық аударма сәйкес келеді (түпнұсқамен бірдей күші бар ресми құжаттың аудармасы).

Бұл жағдайда мәтіндердің баламалылығы заңды түрде (халықаралық құқықтық құжаттар, келісімшарттар, халықаралық ұйымдардың ұсынымдары мен қаулылары және т.б. үшін) немесе ресми түрде расталады. Соңғы жағдайда бірнеше мемлекеттік тілдері бар мемлекеттердің құжаттарын айтсақ, мұндай корпустың классикалық мысалы – Hansard корпусы. Корпусқа канадалық қауымдар палатасының 21,6 миллион ағылшын және 24,1 миллион француз сөзқолданысын [154] қамтитын үш жылдық пікірталас туралы есептер кіреді. Негізінде, параллель мәтіндердің бұл түрі тақырып бойынша шектеулі және әртүрлі лексикографиялық есептерді шешу үшін жеткіліксіз. Сонымен қатар мұндай мәтіндерді тиісті білім салаларының нормативтік сөздіктерін құру немесе нақтылау үшін пайдалануға болады.

Түпнұсқаға авторлық аударма сәйкес келеді (адамның аудармашысы тапсырыс бойынша немесе өз бастамасы негізінде жүзеге асыратын мәтіннің аудармасы).

Бұл жағдай ең көп таралғанымен аудармалардың сәйкестігін қарастыруды қажет етеді. COBUILD жүйесінің конкорданстары болған жағдайда, автор-

лар «тағайындалған» аударманы таңдау үшін жауапкершілікті өз мойнына алады. Параллель корпуста көркем әдебиетті (терминді анықтаудың барлық мүмкін нұсқаларында) және әртүрлі авторлар жасаған аудармалар жиынтығын қолданған жағдайда, әртүрлі аудармалардағы реалияларды, ксенонимдерді, жергілікті, тұрмыстақ элементтерді (лексиканы) аудару нұсқаларын лексикографиялық зерттеу мүмкіндігі пайда болады, демек, тиісті типтегі сөздіктерді де құру мүмкіндігі артады.

Авторлық аударманың тағы бір жағдайы – публицистикалық, ғылыми және ғылыми-техникалық әдебиеттерді аудару. Бұл жағдайда аударманың сәйкестігі туралы мәселесі де маңызды, өйткені мұндай корпусты арнайы сөздіктер жасау үшін пайдалануға болады. Сонымен қатар, мұндай аудармаларды қолданған кезде аудармалардың түпнұсқасына немесе алынған дереккөзге сәйкестігін бағалау керек. Мүмкін, бұл жағдайда редакцияланған және маңызды басылымдарда жарияланған материалдарды қолданған жөн.

Түпнұсқаға машиналық аударма сәйкес келеді (белгілі бір машиналық аударма жүйесімен жасалған аударма).

Жағдайды толығырақ қарастырайық. Мұндай мәтіндік массивтер тілден тәуелсіз онтологияны құру үшін пайдаланылуы мүмкін. Бұл жағдайда онтология дүниеде немесе пән саласында бар ұғымдар, олардың қасиеттері және бір-бірімен байланысы туралы ақпаратты сақтайтын білім қоры болып табылады.

Онтологияның тезаурустан айырмашылығы – оның құрамында тек тілге тәуелсіз ақпарат пен сан алуан мағыналық қатынастар болады, сонымен қатар ол таксономиялық қатынастарды қамтиды. Сонымен, онтологияны құру міндеті – мәтінді мағыналық өңдеуге қажетті дүниенің белгілі бір моделін жасау міндеті. Онтология сөздің лексикадағы мағыналарын білдіретін ұғымдарды анықтауы керек және сәйкесінше:

- ұғымдардың сұрыптамалық (селективті) шектеулерін сақтау керек;
- табиғи тілде және кез келген салада мәтіндерді қайта өңдеу жүйелерінде қолдануға арналған кез келген қолданбалы бағдарламада қолдану мүмкіндігі болуы керек;
- тілден тәуелсіз болуы керек;
- әртүрлі пайдаланушылардың жұмысына ыңғайлы болуы керек.

Әдетте, онтологияны сипаттау үшін кеңейтілетін белгіленім тілі қолданылады (Extensible Markup Language – XML).

Осындай онтологияны құрудың бір нұсқасы корей тілінен жапон тіліне және жапон тілінен корей тіліне машиналық аударма жүйелерін, сондай-ақ Жапонияда құрылған Кадокава [155] тезаурусын пайдаланады, онда 1110 арнайы семантикалық категориялар мен төрт деңгейлі таксономиялық иерархия қолданылады, сонымен қатар семантикалық қатынастардың 30 түрі қосымша бөлінген. Бастапқы онтологияға енгізілген әрбір тұжырымдамада Кадокава тезаурусының коды, корей тіліндегі атауы, ағылшын тіліндегі атауы, тұжырымдаманы енгізу уақытының белгісі және тұжырымдаманың анықтамасы бар.

Ұғымдар кодтарын Кадокава ұғымдарының кодтары арқылы біржақты анықтауға болатынына қарамастан, олардың корей және ағылшын тілдеріндегі

атаулары онтологияны әзірлеушінің оқылуы мен ыңғайлылығы үшін енгізіледі.

Кадокава тезаурусына негізделген онтологияны құру үшін COBALT-J/K және COBALT-K/J екі тілді автоматты сөздіктер қолданылады, олардың көмегімен түпнұсқа және машиналық аударма мәтіндері белгіленеді. Бұл жағдайда атау және етістік сөздер Кадокава тезаурусындағы ұғымдар кодтарымен түсіндіріледі, әр сөздік сөздердің белгілі бір санатының сөздігіне, зат есім сөздігіне, етістік сөздігіне және т. б. сәйкес келетін әртүрлі сөздіктерге енеді, сөздіктерді түзу үшін SELK жобасы қолданылады.

Корей тілінің Седжон (Sejong) [156] электронды лексикасы SELK жүйесі деп аталады.

Осы деректердің негізінде машиналық аударма арқылы онтологияны құру кезінде болат өндірісінің пәндік аймағында патенттік материалдардың параллель корпусы құрылды. Массивтің көлемі тиісті синтаксистік және семантикалық белгіленімдері бар жапон және корей тілдеріндегі 250 мың сөйлемді құрайды, оның негізінде күрделі құрылымдардың аударма нұсқалары мен синтаксистік және семантикалық полисемияларды алу ережелері анықталған.

Мұндай онтологияны

– сыртқы әлем туралы білім көзі ретінде;

– сенімді сөздік ақпарат ретінде;

– полисемияны жою модельдерін құру үшін қажет контекстік ақпарат ретінде қарастыруға болады

Онтологияны құру үшін қарастырылған әдіс мұқият қолмен өңдеуді қажет етеді, яғни синтаксистік қатынастардан адамның нақты ережелері мен түйсігіне сәйкес семантикалық қатынастарға ауысуды көрсетеді.

Параллель мәтіндердің корпусын құруда машиналық аудармаларды қолдану автоматтандырылған лексикография үшін маңызды. Бұл машиналық аударма жүйелерінің лексикографиялық қолдауын құруға қатысты ұсыныстар бойынша сәйкестік сақталған жағдайда терминдердің аудармасының сәйкестігі мен тұрақтылығын бағалауға, сөздіктерді өзгертуге мүмкіндік береді. Сонымен қатар, мұндай параллель корпус белгісіз (анықталмаған) сөздер мен құрылымдарды (конструкцияларды) нақтылауға мүмкіндік береді.

Сонымен қатар мұндай параллель корпус белгілі бір жүйені жөндеудің жұмыс құралы, оқыту үлгісі болып табылатындықтан жүйені өзгерту кезінде ауыстыруды қажет етеді.

Түпнұсқа мен аударма арасындағы теңестірудің экстремалды жағдайы – түпнұсқаның аудармаға емес, басқа тілдегі мазмұны бойынша теңестірілетін мәтінге сәйкес келуі.

Бұл жалған параллель (псевдопараллель) деп аталатын мәтіндер жаңа терминдердің сөздіктерін құрастыру үшін пайдаланылуы мүмкін, бірақ мұндай дереккөздерден жасалған мәтіндік корпус басқа (ерекшеленетін) белгіленім түрімен қолданылуы керек.

Мәтіндік корпусты белгілі бір мақсатта пайдалану мүмкіндігі туралы мәселені шешу үшін түпнұсқа-аударма қатынасы маңызды рөл атқарады. Сонымен қатар оны қолданыстағы корпусты пайдаланатын немесе жаңасын жасайтын әзірлемешілер білікті түрде бағалап жасауы керек.

Сондықтан параллель мәтіндер корпусын құру міндеті үшін, жалпы корпус лингвистикасы үшін құрылатын корпусстарды жасаушылар (әзірлеушілер) мен пайдаланушылардың біліктілік аспектісі өте маңызды. Шынында да, кез келген мәтінді электронды форматта дайындауға жалпы көшу, бір жағынан, дербес компьютерлер мен Интернет желісінің қол жетімділігі, екінші жағынан, корпусық лингвистиканың дамуына алып келді. Яғни мәтіндердің шексіз массивтерін зерттеуге негізделген лингвистикалық бағыт пайда болды. Мұндай мәтіндерді сақтау, толықтыру және зерттеу мүмкіндігі жазбаша аударма тәжірибесінде жаңа әдістің – аударма жады бар жүйелердің пайда болуына себепкер болды.

Бұл массивтер бір тілдегі зерттеулерге (бір функционалды стиль, бір тақырып, белгілі бір автордың шығармалары және т.б.) және әртүрлі тілдердегі параллель немесе шартты параллель мәтіндерді карама-қарсы зерттеуге мүлдем жаңа мүмкіндіктер береді.

Автоматтандырылған лексикография осы фонда мүлдем жаңа болып көрінеді, ол кезде корпусы пайдаланушы нақты қолданылатын сөздік қорды таңдай алады және аударманы немесе мағынаның ауқымын үлкен массивте сипаттай алады, мұны қолмен емес, компьютердің көмегімен жасайды. Осы бағытта бұрыннан бар немесе жасалған мәтіндік корпусы талдау және жаңаларын құру міндеттері ерекше орын алады. Арнайы міндет – лексикалық білімді алудың әдістері мен құралдарын әзірлеу (мәтіндегі лексикалық бірлікті тану, сөз – сөз тіркесі – аударма сияқты үштікті (триада) қалыптастыру), білімді берудің әдістері мен құралдарын әзірлеу үшін параллель мәтіндерді пайдалану, арнайы сөздіктер жасау, көптілді конкорданстарды құру.

Толық мәтінді деректер қорының, сондай-ақ параллель мәтіндердің теңестірілген корпусының көп болуы лингвистикалық автомат (LA) орындайтын операциялар шеңберінде жаңа қосымша ішкі жүйенің пайда болуына әкелді – автоматтандырылған жүйеде әртүрлі типтегі сөздіктер жасаудың ішкі жүйесі. Бұл ретте жиілікті, әліпбилі, кері сөздіктер, конкорданстарды дайындау тапсырмалары әдеттегі арнайы алгоритмдермен жүзеге асады. Жаңа аспект – екі тілді сөз қорын (резиденттік сөздіктер немесе компьютердегі жай сөздіктер) жасау және қолдау, сондай-ақ машиналық аударма жүйелері үшін автоматтандырылған сөздіктерді жасау және/немесе жаңарту үшін LA пайдалану.

Бұл мәселелерді шешу көбінесе екі тілді параллель мәтіндерді теңестіру әдістерімен анықталады. Параллель корпуста контексті зерттеудің статистикалық әдістерін қолдану сөздің полисемиясын алып тастауға немесе азайтуға және аударма тілінде оның баламасын таңдауға мүмкіндік береді [157].

6 ОҚУ КОРПУСЫ

6.1 Қазақ тілінің оқу корпусын әзірлеу мәселелері

Ұлттық корпустарды құрастыру ісі қазіргі цифрлық заманның кезек күттірмейтін аса маңызды әрі өзекті мәселесі болып отыр. Бұл ретте корпустар жасау ісі әлемдік лингвистикада өткен ғасырдың орта шенінде-ақ басталып кетті. Алғашқы ағылшын тілінің корпусын жасау тәжірибесі басқа да еуропа тілдерінің, орыс, одан қала берді түркі тілдерінің корпустарын жасауға негіз болды. Бұрын қол жұмыстары арқылы орындалған зерттеу және оқыту ісі корпустарды пайдаланудың арқасында үлкен нәтижелерге қол жеткізді. Корпустар пайдаланушыны мол материалдармен қамтамасыз етті, жаңадан сөздіктер құрастырып, лингвистикалық зерттеулер жүргізуді жеңілдетті. Корпусты пайдаланушылар саны күннен-күнге артты. Мұндай қажеттіліктер корпустарды пайдаланушы мақсатына қарай түрлендіріп жасауға ықпал етті. Тарихшылар немесе тарихи тіл білімін зерттеушілер үшін – Тарихи корпусты, аудармашылар мен тіл үйренушілер үшін – Параллель корпусты, сөйлеу ерекшеліктерін зерттеушілер мен сөз мәдениетін дамыту үшін – Ауызша корпустарды жасау қажеттілігі туды. Осы ретте білім беру ісін цифрландырудың бірден-бір құралы болатын – Оқыту, білім беру корпустарын жасау да аса қажеттілік туғызды. Оқыту корпусы дегеніміз – тіл үйренушілерге арналған ақпараттық-корпустық база.

Бүкіл тіл жүйесін біртіндеп компьютер жадына енгізіп, оларды белгілі бір автоматты басқару орталығы арқылы іске қосу, әсіресе, тіл тұтынушыларының практикалық қажеттілігі үшін аса маңызды. Бұрын ғылыми зерттеулер мысал жинау, кәргішкі жазу, оларды жинақтау, топтау сияқты қол жұмыстары арқылы атқарылып, ұзақ уақытқа созылған ғылыми ізденісті қажет еткен болса, тілдік корпустар мұндай жұмыстарды барынша жеңілдететіндігі және де тілші мамандар үшін ғана емес, журналистер, әдебиетшілер, тарихшылар, мәдениеттанушылар, тіпті қалың көпшілік қауым үшін де ақпараттық-анықтамалық құрал қызметін атқару мүмкіндігі бар екендігі анықталып отыр. Осы орайда білім беру жүйесін автоматтандыру ісі қазіргі педагогикада аса өзекті мәселелердің біріне айналды. Бұл өз кезегінде мұғалімдерді шығармашылық әрекеттен, қол жұмысынан босатып, білім беру жүйесін қарапайым механикалық амалдарға ауыстырады. Білім беру корпусы жасалса, білім беру барысындағы дәстүрлі әдіс-тәсілдерге жұмсалатын адам еңбегі компьютерге (білім беру корпусына) жүктеледі. Бұл адам баласының бір рет ғана атқаратын еңбегін ізденушілер өз қалауынша компьютердің көмегімен сан рет қайталауына мүмкіндік алады [158, 593].

Білім беру/оқу корпусы – оқытуды дараландыру үшін компьютерлік негіздегі техникалық, оқу-әдістемелік, лингвистикалық, бағдарламалық пен ұйымдастырушылықты қамтамасыз етуге арналған ішкорпус. Корпус арқылы білім беруді «автоматтанған оқу жүйесі» құрайды [158, 593].

Білім беру корпусындағы деректер: 1) тіл үйрену; 2) орфографиялық, грамматикалық, синтаксистік ережелерді меңгеру; 3) оқулықтар, оқу құралдарын,

эртүрлі жаттығулар жинағын құрастыру; 4) онлайн мультимедиалық оқыту сервистерін құру; 5) ғылыми лингвистикалық зерттеулер жүргізу үшін аса пайдалы.

Шет тілін (ағылшын) үйренушілердің сөйлеу мәтіндерінің корпусын жасау өткен ғасырдың 90-жылдары басталған. Мәселен, осындай корпусардың бірінде 10 информанттың ағылшынша сөйлеген сөзінен жазып алынған қате қолданыстар іріктелген 2000 сөзқолданыстан тұратын картотекалық қор бар. Осындай қате қолданыстар негізінде типтік қателерді анықтап, «өзекті сөздік», «өзекті грамматика» тұрғысынан тіл үйренуші лексикасы мен грамматикасындағы қателіктерді түсіндіретін, жөндейтін қосымшалар жасаған, белгіленімдер енгізген. Мұндай тіл үйренушілер сөйлеуінен жинақталған корпусар материалдарын тілдің төл иелмендері мен тіл үйренушілердің сөйлеуін салыстырып, эртүрлі қорытындылар жасаған. Осы сияқты корпус мысалы, Эли Хинкель корпусы (500 мыңдай сөзқолданыстан тұрады) [159].

Ең танымалы: «International Corpus of Learnt English» атты халықаралық ағылшын корпусы (1990) Бельгияда жасалған [160]. Бұл корпусқа ағылшын тілін жетік меңгерген студенттердің эсселері енгізілген.

Тағы бір танымал оқу корпусы С.Гранже корпусында 14 тіл ұстармандарының эсселер жинағы берілген [161].

Гонконг корпусында ағылшын тілін үйренуші студенттердің жазба жұмыстары жинақталған т.б.

Оқу корпусарын талдаған зерттеушілердің көрсетуінше, қазіргі Оқу корпусарының көпшілігінің мәтіндік базасы 50 000-150 000 сөзқолданыстан тұрады. Алайда көлемді Оқу корпусары жоқ емес. Мысалы, the Quebec learner corpus-ында 250 000-нан аса сөзқолданыс [162].

Орыс тілінің оқу корпусы 2007 жылы қолданысқа енгізілген. Бұл корпус орыс тілі мен әдебиетін мектепте оқытуға арналып жасалған. Бұл корпусы сонымен қатар жоғары оқу орындарындағы білім беру кезінде қолдануға болатыны айтылады. Әсіресе, өзге салалар мамандарына тіл үйретуде, өзге ұлт өкілдеріне тіл меңгертуде қолданады [39]. Орыс тілінде оқу корпусары негізінде оқушылар жіберетін типтік грамматикалық, стилистикалық қателіктер туралы да зерттеулер жазылған [163]. Оқыту корпусын кейде «Корпусы ошибок» деп те атайды [164].

Қазақ тілінде Білім беру корпусын Ұлттық корпусының құрамдас бөлігі ретінде әзірлеу қажеттігі және оның мәтіндік базасы мен белгіленімдері туралы алғаш айтқан ғалым – профессор Асқар Құдайбергенұлы Жұбанов болды. А.Жұбанов Білім беру корпусын құрастыруда келесі міндеттерді негізге алу қажеттігін айтады: – Білім беру корпусына енетін мәтіндердің метамәтіндік параметрлері компьютерлік базаға салынуы қажет. Сондықтан корпус құрастыруда мета белгіленімдер (оқулық атауы, авторы, жылы, шыққан орны, стилі, жанры, көлемі т.б.) мен грамматикалық белгіленімдерді мәтінге енгізудің әдіс-тәсілдерін, түрлерін, компьютерлік бағдарламаларын т.б. зерттеудің маңызы зор; – Эртүрлі сөздіктер құрастыруда, ғылыми грамматикалар жазуда, оқыту жүйесінде және оқулықтар құрастыруда, әдістемелік құралдар жазуда, аударма жұмыстарында, ғылыми зерттеулер жүргізуде, статистикалық мәліметтер

алуда, энциклопедиялық ақпараттар алуда осы білім беру корпусын негізге алу керек; – Білім беру корпусына енген мәтіндердің құрамына тек қазақ тілін ғана емес, басқа да пәндер қосу және мәтіндерде кездесетін омоним сөздерді автоматты түрде немесе жартылай автоматты түрде ажырату қажет [12].

Қазақ тілінде Оқу корпусын әзірлеу туралы мәселе А.Байтұрсынұлы атындағы Тіл білімі институтында А.М.Фазылжанның жетекшілігімен **«Қазақ мәтінін автоматты тану: лингвистикалық модульдер мен IT-шешімдер әзірлемесі»** атты бағдарламалық-нысаналы қаржыландыру бойынша зерттеу жобасының аясында (өтінімнің 2-бөлімі) арнайы көтеріліп, зерттеліп, жасақталып жатыр.

Оқу корпустарын жасаудың көпшілігінің басты мақсаты – оқытылып отырған тілді меңгерудің әдіс-тәсілдерін анықтау мен оның тиімділігін талдауға бағытталған. Әсіресе мектеп оқушылары мен өзге тілді адамдарға тіл меңгеру ісін оңтайландыруды көздейді. Тіл үйрену кезіндегі жазба жұмыстары мен ауызша сөйлесімдер базасын жасақтау арқылы, корпустық мәтіндік база жинап, ондағы типтік қателерді анықтау арқылы тіл үйрету процесіндегі маңызды, назар аударуға тиісті нәрселерді шығарады. Сөйтіп, тіл меңгеру, білім беру бағдарламалары мен әдіс-тәсілдеріндегі кемшіліктерді анықтауға мүмкіндік береді.

Оқу корпустарын құрастыруда тіл меңгерудің деңгейлеріне лайықтап жасақтау мәселесі де басты назарда болуы керек. Өйткені тілді мүлде білмейтін басқа ұлт өкілі мен бұл тілде сөйлей алатын тілүйренушінің қажеттілігі бірдей емес. Осы орайда оқу корпустарын құрастырушылар кейде оны тіл деңгейлеріне қарай бөліп жасақтайды.

Қазақ тілінің Оқу корпусын әзірлеуде де деңгейлік оқыту принципін ұстану қажет деп есептейміз. Айталық, қазақ тілін оқытуда жоғарғы оқу орындарында «Практикалық қазақ тілі», «Кәсіби қазақ тілі» пәндері бар. Бұл пәннің алғашқысы орыс аудиториясында оқитын 1-курс студенттеріне оқытылады. Ал екіншісі мамандықтары бойынша терминдерді меңгеру мен кәсіби сөйлеуге бағытталады. Сол сияқты қазіргі уақытта жоғарғы оқу орындарының Филология факультеттерінде «Шешендік өнер» т.б. атаулармен қазақ тілінде шешен сөйлеуге бағытталған арнайы элективті курстар жүргізіледі. Ісқағаздарын жүргізу бойынша да пәндер оқытылады. Қазақ мектептерінде оқытылатын қазақ тілі пәні орта мектеп шеңберіндегі қазақ тілінің ғылыми аппаратын игеруге бағытталуы тиіс. Осы тілдік деңгейлерге қарай Қазақ тілінің оқу корпусының мәтіндік базасы мен лингвистикалық құралдары төмендегі ретпен жасақталуы тиіс:

- *Практикалық қазақ тілі;*
- *Кәсіби қазақ тілі;*
- *Теориялық қазақ тілі;*
- *Іскери қазақ тілі;*
- *Шешендік қазақ тілі.*

Орыс тілінің оқу корпусының **мәтіндік базасы** бірнеше стильдер мен жанрларды қамтиды. Корпустың Іздеу жүйесі үлкен екі ұяшыққа жинақталып берілген:

- 1) Көркем мәтіндер;
- 2) Бейкөркем (нехудожественные) мәтіндер.

Көркем мәтіндер негізінен жоғары сынып оқушыларының қазақ тілі мен әдебиет хрестоматияларында берілетін және сыныптан тыс оқу ұсынылатын көркем шығармалардан жинақталған. Біздің түсінуімізше, орыстың классикалық жазушылар шығармаларын алған. Ал бейкөркем мәтіндер орыс тілі мен сөз мәдениеті пәндерінде оқытылатын басқа да функционалды стильдерден алынған. Мысалы:

- Публицистика;
- Ресми-іскери;
- Ғылыми (мектеп);
- Сөйлеу.

Жанрлық құрамы жан-жақты, негізінен мектеп бағдарламасы аясындағы мәтіндерге негізделген. Атап айтқанда, *мақала, сұхбат, күнделік, өтініш, іскери хат, ғылыми не ғылыми-көпшілік мақала, сараптама, аңдатпа, реферат, тұрмыстық хаттар, күнделікті сөйлеу т.б.*

Қазақ тілінің оқу корпусын жасауға қатысты А.Жұбанов «Білім беру корпусына – ең алдымен мектеп бағдарламасы бойынша оқытылатын оқулықтар және ЖОО студенттерінің пән бойынша оқытылатын шығармалары кіреді», – деген болатын.

Қазақ тілінің оқу корпусының мәтіндік базасына қандай мәтіндер енгіземіз?

– Көркем шығармалар – *Жалпы тіл үйренушілердің бәріне (1) әдетте мектеп оқулықтарының хрестоматияларында берілетін классикалық шығармалар, 2) балалар әдебиеті; фольклор (ертегі, жыр, жұмбақ, жаңылтпаш, өлеңдер, мақал т.б.);*

– Тақырыптық мәтіндер (*практикалық қазақ тілін үйренушілер үшін (Отбасы, Денсаулық, Саяжай, Саяхат, Дүкенде т.б.)*)

– Ғылыми мәтіндер – *Кәсіби қазақ тілін меңгерушілер үшін (басқа пәндер бойынша)*

– Публицистикалық мәтіндер – *Газет-журналдарға мақала жаза алу үшін («Балдырған», т.б.)*

– Ісқағаз мәтіндері – *Ісқағаздарын жазып үйрену үшін (өтініш, хабарландыру, келісімшарт, құттықтау, қызметтік хат т.б. үлгілері)*

Жалпы шетелдік оқу корпустарында, жоғарыда айтып өттік, типтік қателерді анықтау үшін, мектеп оқушыларының эсселері, т.б. жазба жұмыстарын корпус базасына салған. Алдымен жаппай жазба жұмыстарын жинақтап, эксперименттер жүргізген. Екінші кезекте оқушылар жазбаларының қатесін корпустық менеджерлер арқылы тексеріп, типтік қателерді анықтаған. Оқу корпустарын компьютерлік қамтамасыз ету шет тілін меңгерудің тиімділігін арттыруға әсер етеді. Өйткені корпустық менеджерлер (құрылғылар) орындайтын мүмкіндіктерді қол жұмысы арқылы жүзеге асыру өте қиынға соғады. Бұл туралы М.С.Мальцева: «Как правила, учебный корпус отражает результаты типичной деятельности, представляя одинаковые письменные задания, выполненные обучаемыми их эквивалентных групп. Учебный корпус позволяет

анализировать продукты учебной деятельности студентов, выявляя их успеваемость, наиболее типичные лексические, стилистические и грамматические ошибки, проблемы в знаниях, на которые следует направить особое внимание преподавателям в дальнейшей работе с обучаемыми», – деп тұжырымдайды [162].

Мұндай типтік қателер орыс тілін оқытудағы бағдарламалар мен оқыту әдістеріне өзгерістер жасауға ықпал еткен. Сөйтіп, орыс тілінде тіл үйрету мәселесін шешуде Оқу корпустарының маңызы зор.

Ал енді қазақ тілінде оқу корпустарының осы қызметін қалай жүзеге асырамыз деген сұрақ туындайды. Бұл ретте орыс тілінің оқыту корпустарындағыдай эсселер базасын жинақтау мен ондағы типтік қате қолданыстарды анықтау үшін біздің ойымызша:

Мектеп платформаларындағы оқушы жазбалары, эсселер (мектептерден алу) (Күнделік кз)

– *Қазақ тілінен емтихан платформаларындағы эсселер (жоғары оқу орындарынан алу)*

– *Арнайы эссе жаздыру (мектеп мұғалімдерімен байланыс орнату) (эксперимент) арқылы электронды жазба жұмыстарын жинақтауға болар еді.*

Мұндай типтік қателерді О.А.Фирсова бірнеше түрге бөледі:

– Лексикалық қателер;

– Грамматикалық қателер;

– Стилистикалық қателер.

Лексикалық қателерге дұрыс емес сөзқолданыстарын, сөздердің тіркесіміндегі қате тіркестерді, сөз мағынасын дұрыс түсінбеу сияқтыларды жатқызады. *Грамматикалық қателер* – род, число, падеждегі, олардың байланысуындағы, синтаксистік құрылымдардағы қателер. *Стилистикалық қателер* деп функционалды стиль түрлерін дұрыс, орынды қолданбауды атаған, яғни мәселен, ғылыми мәтіндерде ауызекі сөйлеу тілінің элементтерін қолдану т.б.

Қазақ тілінің Оқу корпусын жасау егер басқа шетелдік оқу корпустарындай сауаттылыққа бағытталса, онда оның жұмысын мынадай корпустық қосымшалар арқылы реттеуге болады:

– *Мәтіндік редактор;*

– *«Сөзтексер» қосымшасы;*

– *Тіркесімділік сөздігі;*

– *Жиілік шығару программасы*

Оқу корпустарындағы типтік қателерді анықтау мәселесімен байланысты екінші бір қажеттілік – оқушылар, оқытушылар, тіл үйренушілер үшін Сауаттылық құралдарын Оқу корпусының жұмыс үстеліне енгізу болып табылады.

– *Жазу сауаттылығы (Орфографиялық сөздік, емле, пунктуациялық анықтағыш)*

– *Оқу сауаттылығы (Орфоэпиялық сөздік, орфоэпиялық конвертор)*

– *Лексика-грамматикалық сауаттылық (Оқулықтар, түсіндірме сөздіктер т.б.)*

Лингвистикалық белгіленімдердің ішінде корпустағы грамматикалық белгіленім Оқу корпустарында маңызды рөл атқарады. Оның ішінде морфоло-

гиялық белгіленімдер Оқу корпустарының негізгі белгіленімі болып келеді. Оқу корпустарындағы морфологиялық белгіленімдер мектеп бағдарламасына негізделген дәстүрлі грамматикаға сүйенеді. Оқу корпустарында Академиялық грамматика материалдары әлдеқайда жеңілдетіліп берілген, күрделі, даулы мәселелер алынып тасталған. Бұл жайлы ғалымдар: «Морфологическая разметка в обучающем корпусе адаптирована к стандартным нормативным пособиям по русскому языку и обеспечивает именно традиционный, упрощенный грамматический разбор», – деп көрсетеді [162].

Орыс тілінің мектеп бағдарламасына бағытталған белгіленімі бар, **омонимдер қатары ажыратылған корпус** – орыс тілінің оқыту корпусы. Аз көлемде болса да оқыту корпусына алынған мәтіндерді омонимдерін ажыратып берген, яғни оқушы шатаспау үшін, Оқу корпусын пайдаланушылар үшін омоним сөздердің қай сөз табына жататынын, қандай грамматикалық форма екенін нақты көрсеткен. Алайда омонимдер қолмен ажыратылған. Өйткені автоматты әдіс кейбір қате талдауға ұрындыруы мүмкін. «В корпусе представлены тексты **со снятой вручную омонимией**» – деп мұны корпус әзірлемешілері атап көрсетеді.

А.Жұбанов корпустағы омоним мәселесі туралы: «Табиғи тілді талдау автоматты түрде айнақатесіз жүргізіледі деуге болмайды, себебі, ол, әдетте, бір лексикалық бірліктің (сөз, сөзтіркес, сөйлем) бірнеше талдау нұсқасын ұсынады. Мұны *грамматикалық омонимия* деп атайды. Осындай проблемалық мәселелерді (морфологиялық, синтаксистік) жою компьютерлік лингвистика саласының ең бір маңызды және күрделі мәселелеріне жатады», – деп көрсеткен болатын.

Қазақ тілінің Оқыту корпусына салынған мәтіндерге түгелдей морфологиялық белгіленімді автоматты программа арқылы жүзеге асыру мүмкін емес. Өйткені тілдегі омоним, омограф, омоформалар бір сөз туралы бірнеше нұсқаны, яғни омонимдер қатарын көрсетеді. Тіл үйренуші тілді толық меңгермеген жағдайда, шатасуы мүмкін. Сондықтан автоматты морфологиялық талдаулар жасау бағдарламасына қосымша, омонимдерді қолмен ажырату қажеттігі туындайды. Мәселен, төмендегі омоним, омоформаларды компьютерлік бағдарлама автоматты түрде қай мағынада тұрғанын ажырата алмайды.

Қара жол
Бері қара
Қара танымайды

Сал боп ауырды
Көре сал
Сал адам

Көр соқыр
Көр қазу
Кино көр

Біздің ойымызша, оқу корпусына салынатын мәтіндердегі омонимдерді түгелдей ажырату мүмкін емес. Шетелдік корпустарда осы себепті оның көлемі

аз боп келеді. Бұл ретте Оқу корпусына мәтіндерді *белгіленген* және *белгіленбеген* мәтіндер деп екіге бөліп орналастыру қажет. Олай болмаса, белгіленген мәтіндер көлемі аз болып, Оқу корпусының мәтіндік базасы азайып қалады.

Қазақ тілінің оқу корпусы кімдерге арналып жасалады?

- Оқытушыларға (әртүрлі деңгейде қазақ тілі пәнін өткізу кезінде);
- Оқулық құрастырушыларға;
- Екінші тіл ретінде қазақ тілін оқушыларға (Практикалық қазақ тілі);
- Кәсіби тілдесім мен кәсіби терминологияны меңгерушілерге (Кәсіби қазақ тілі);
- Қазақ тілін ғылым тілі ретінде оқитын оқушылар мен студенттерге (Теориялық қазақ тілі);
- Қазақ тілінің дәстүрлі сөзсаптамын жатық меңгергісі келетіндерге (Шешендік қазақ тілі);
- Жалпы тіл (қазақ тілін) үйренушілерге.

Орыс тілінің оқыту корпусында «проект-спутники» деп аталатын қосымша жұмыс істейді. Бұл қосымша корпус материалына негізделген жаттығулардан, мектеп платформаларындағы әдістемелік құралдардан тұрады.

Автоматтанған оқу жүйесі әртүрлі пәндерді оқыту барысында пайдаланылуы мүмкін. Оның қызметі сол жүйедегі программалардың оқушыға тигізетін әсері арқылы анықталады. Оқыту программаларының негізгі қызметтеріне мыналар жатады: - *Жаттығу - Бақылау - Бағалау* [158, 130]. Бұлар Білім беру корпусына тікелей қатысты программалар. Жаттығу программасының негізінде мұғалім сабақ үстінде корпустан қажетті иллюстративті материалды іздей алады немесе жаттығуларды құрастыра алады. Немесе ЖОО-да студенттер семинарда пайда болған нақты қойылған мәселелерді шешуде сұрақтарға жауап бере алады. Мұғалім немесе оқушы, студент корпустан сөздің белгілі бір формасын іздеп-тауып, солардың негізінде неше түрлі жаттығулар құрастыруына болады. Ал бақылаушы программалардың жұмысы кезінде әрбір тапсырманың жауаптары базада сақталып отыруы қажет. Жауаптардың негізінде кеткен қателердің ақиқат және кездейсоқ екенін ажырата алады. Бағалау программаларында қателерді әрі қарай талдап, тапсырма нәтижелерін статистикалық өңдеуден өткізеді. Осылай корпустен жұмыс істеуде студенттің, оқушының шығармашылық қабілеті ашылады. Мәселен, қолданушы корпустан сөздің белгілі бір формасын іздеп-тауып, солардың негізінде неше түрлі жаттығулар құрастыруына, мысалдар жинақтауына болады. Тапсырмаларды өз бетімен де, үй жұмысы ретінде де орындай алады.

Қорыта келгенде, Қазақ тілінің оқу корпусы лингвистикалық және лингводидактикалық талдаулар үшін репрезентативті база болып табылады. Қазақ тілін меңгеру процесінің ақиқат бейнесін көрсетеді. Тіл үйренудегі бақылау, өздік жұмыстарды әлдеқайда жеңілдетеді. Оқушылардың жазба жұмыстарын, атап айтқанда, рефераттар жазу, зерттеулер жүргізу сияқты қажеттіліктерін өтейді.

6.2 Қазақ тілінің лексикалық қабатын оқу корпусына бейімдеу: семантизация және автоматтандыру жолдары

Оқу корпусы (learner corpus) – тіл үйренушілердің деңгейіне сай оқытатын мәтіндердің электронды қоры ғана емес, әрбір сөздің лексикалық қабатына тән сипатын танытатын тілтанымдық жүйе. Сондықтан оқу корпусы тілтанымдық әрі лингводидактикалық қызметті қатар атқарады. Оқу корпусы – тіл үйренушілерге қосымша құрал ретінде, әсіресе қате қолданысын түзетуге көмекші болатын автоматтандырылған сөздік қор. Оқу корпусын әзірлеу және оны арнайы зерттеуге деген қызығушылықтың себебі қоғам сұранысынан туындап отыр. Қазіргі қазақ қоғамында мемлекеттік тілді үйренушілер ғана емес, қазақы тілдік ортаның өзінде ана тілінде дұрыс жазу, дұрыс сөйлеу, дұрыс сөйлем құрастыру, әрбір сөзді мағынасына сай дұрыс қолдануға деген талпыныс оянды. Осы тұрғыдан қоғамның жазарман қауымы ғана емес, жалпы халық қазақ тілінің саламатты қолданысы мен дамуына жанашырлықпен қарайтынын байқадық. Себебі, жалпыхалықтық сұранымда сөз мағыналарының түсіндірмесі, жаңа сөздердің аудармасы, әдеби тіл мен жергілікті немесе сөйлеу тілдерінің қолданыс ортасы мен мәртебесі, қатесіз жазу және қатесіз сөйлеудің нормалары, сөздердің антонимдік, синонимдік, омонимдік сипаты, бейнелі қолданысы және фразеологиялық қордың мағынасы өзге тілдегі баламасы, т.б. басымдыққа ие болып отыр. Тіл тұтынушылардың тілтанымдық сұранымдарын өтейтін қолжетімді әрі оңтайлы лингвистикалық құрал – оқу корпусы болмақ. Оқу корпусы қоғам мұқтажындағы тілтанымдық лексикалық және нормативті қордың электронды жиынтығы ретінде ұсынылатын ресурс. Бұл тақырыпта оқу корпусының шетелдік және отандық лингвистикадағы зерттелуіне, негізгі құрылымдық жүйесіне, браузерлерге орналасатын лингвистикалық контентті әзірлеу тәсілдеріне тоқталамыз.

Қазақ тілінің лексикалық қабатын оқу корпусына жүйелі енгізу үшін қазақ тілінің нормативті, екі тілді және түсіндірме сөздіктері материалдық база ретінде алынады. Оларды біз үш топқа жіктедік:

1. Материалдық қор – нормативті қор. Оған орфографиялық және орфоэпиялық сөздік алынды. Оқу корпусына нормативті сөздік түгел енгізілмейді. Қателер жиілігіне қарай сұрыпталған, яғни босаң нормадағы немесе жазарман жиі қателесетін сөздер іріктеп алынады. Бұл жағдаятты анықтау үшін жинақтау әдісі қолданылды. Мұғалімдердің көмекші құралы болатын сайттар мен әлеуметтік желілердегі сұрақ-жауап айдарындағы қателерге байланысты жиі сұралатын сөздер тізбегі жасалды. Содан соң бастауыш және қазақ тілі мұғалімдерінен сұхбаттасу немесе сауалнама әдісі арқылы оқушылардың жиі жіберететін қателері анықталады. Осы зерттеулерді жинақтай келе, эмпирикалық зерттеу әдісі арқылы емлесі қиын сөздер тізбегі құрылады. Барлық зерттеулердің нәтижесінде жиналған сөздер тізбегінің прагматикасы, яғни іздеушіге ыңғайлылығы ойластырылады.

2. Материалдық қор – лексикалық қор. Оқу корпусының бұл қызметі лексикалық жүйеге негізделеді: төл/кірме сөз; көне/жаңа сөз; әдеби/бейәдеби (сөйлеу, бейпіл,) сөз; термин/жергілікті, кәсіби; тарихи сөз;

3. Материалдық қор – семантикалық қор. Бұл қызмет семантикалық жүйеге негізделеді.

Омонимдер, антонимдер, синонимдер, полисемия, эврисемия тізбегі, фразеологиялық және бейнелі сөздер тізбегі.

4. Материалдық қор – тіл үйрену деңгейіне сәйкес үш тілді лексика-семантикалық минимум. Оған сөздер тізбегі (әдеби сөз, терминдер, фразеологизмдер, мақал-мәтелдер, т.б.) және мәтіндер желісі енеді.

Оқу корпусының әзірлемесін дайындау барысында бірқатар нұсқалар болған. Оқу корпусына мәтіндер желісі жеткіліксіз болды. Сондықтан лексика-грамматикалық минимумға сүйенуге тура келді. Бұл тұжырымдар шетелдік ғалымдардың тәжірибесінен туындап отыр. Алғашқы оқу корпусы – Кембридж оқу корпусы (Cambridge Learner Corpus) деп саналады. Артынша ICLE (Ағылшын тілін үйренуге арналған халықаралық корпус), LLC (Лонгманның оқу корпусы) іске қосылды.

Оқу корпусының құрылымында іздеу браузерлері көп болған сайын рейтингісі жоғары, яғни қызмет көрсетуіне сұранысы көп корпуста айналады. 1990 жылы Азия мен Еуропада қызмет көрсететін оқу корпустарының ішінде Белгияда құрылған оқу корпусының құрылысы ыңғайлы деп танылды. Бұл – халықаралық ағылшын тілді ICLE корпусы (International corpus of Learner English). Онда – жаңа мәтіндер, яғни халықаралық қатынастағы үстем тілдерде жазылған студенттердің жаңа танымындағы эсселері енгізілді [165]

Кейінгі құрастырылған оқу корпусы салғастырмалы вокабуляр оқулықтарға негізделді. Яғни екі тілді дискурс, тілдескіш, екі тілді антонимдер сөздігі, т.б. тіл үйренушілердің сұранысына қарай құрастырылды. Осы корпус-тың қатарына ауқымы кең С. Гранженің оқу корпусы енеді. Гранже корпусы 14 ұлттық тілді тұтынушылардың эсселер жиынтығынан тұрады. Гонконг оқу корпусы ағылшын тілін үйренуші студенттердің жазба жұмыстарына басымдық береді. Тіл үйретуді мақсат ететін оқу корпустарына Мичиган корпусы (Michigan Corpus of Academic Spoken English (MICASE)), Тампер корпусы (ELFA (English as Lingua Franca in Academic Settings)) жағады. Аталмыш электронды тіл құралдарын әзірлеуші Анна Мауранен [166].

О.Н.Кашмилованың зерттеуі бойынша: «Өзге тілді үйрететін оқу корпусының міндеті тіл деңгейіне, тақырыпқа ыңғайлы мәтіндер мен сөздерді тауып беру ғана емес. Мәтін корпусымен жұмыс істеудегі құралдар (жиілік тізім, конкорданс, парсер) лексикалық қабаттағы қателерді анықтаумен қатар, тіл үйренушілер үшін маңызды сөздіктерді құрауға, грамматиканың өзекті мәселелерін анықтауға, шешілмеген тіл заңдылықтарын тануға мүмкіндік береді. Себебі оқу корпусы тәжірибедегі шынайы үдеріске қызмет ететін тілтанымдық құрал» [167].

С. Гранже мен Н. Нессельхауфтың пікірінше, оқу корпусын құру критерийлері, біріншіден, құрастырушының зерттеу мақсаттарына, екіншіден, деректер жинағының шектеулеріне байланысты. Олардың ұсынған оқу корпусы 7 ұяшыққа ажыратылған:

- 1) оқушылар корпусы;
- 2) ғылыми-техникалық мәтіндер корпусы;

- 3) әдістемелік мәтіндер корпусы;
- 4) көркем мәтіндер корпусы;
- 5) ауызша диалог корпустары;
- 6) жеке хаттар корпусы;
- 7) журнал мақалаларының корпусы [161].

Е.П.Соснованың пікірінше: «Оқу корпусын ұйымдастырудың негізгі көзі оқытылатын тілді игерудің әдістері мен тиімділігін анықтау үшін оларды талдауға негіз болатын әрі дәл келетін материалдар жиынтығы болып табылады» [168].

В.В. Рыкова: «Оқу корпусы тіл үйренушілердің жазба жұмыстарындағы қатені көрсету үшін, сөздік қорын, көп қолданатын сөздер жиілігін көру үшін, сөзқолданысы қай стильге жататынын білу үшін аса маңызды», – дейді [169].

Зерттеушілердің тұжырымын сараптай келе, оқу корпусы біріншіден, оқу сауаттылығына үйретуші құрал, екіншіден, тіл үйренуші электронды қосымша, үшіншіден – тіл деңгейіне лайықталған мәтіндік қор деген қорытындыға келдік. Осы зерттеулерді негізге ала отырып, қазақ тілінің лексикалық қабатын оқу корпусына бейімдеу жолдарын талқылап, нәтижесін ұсынайық.

Ең алдымен қазақ тілінің оқу корпусына қандай мәліметтер салыну керек? Қазақ тілінің корпусының мақсаты қандай болмақ? Қазақ тілін үйрену үшін қандай тілдік құралдар қажет? деген мәселелерге шолу жасау нәтижесінде мынаны анықтадық:

– Қазақ тілінің оқу корпусы 2 мақсатты жүзеге асыруы қажет. Біріншісі, қазақтілді азаматтарға тілтанымдық сауаттылықты одан әрі жетілдіру үшін, яғни күмәнданған сөздің дұрысын көретін, тілтанымдық ережелерден қарауға оңтайлы әрі қолжетімді мүмкіндік жасайтын іздеу құрылғысын әзірлеу. Екіншісі, қазақ тілін үйренушілерге арналған электронды қосымша. Оған тіл деңгейіне қарай, тақырып деңгейіне қарай сөздік, тілдескіш, бейне, аудио көрнекіліктер, мәтіндер (топик) тізбегі енеді.

Мақаланың зерттеу нысанына қазақтілді ортаға арналған лексикалық сауаттылықты жетілдіру қоры алынып отыр. Қазақтілді ортада сөз мағынасының ауысып, орынсыз қолданылуын түзетпегеннен қисынсыз тіркестер, жүйесіз сөйлемдер туындап жатыр. Мәселен, қазіргі қазақ қоғамында үлкені де кішісі де «қуып кетті» (не болса соны айтып, әңгімеден ауытқу, бос сөйлеу, көпіру мағынасындағы) тіркесті жиі қолданады. Әрине, бейәдеби сөздер қатарына тіркелген сленг деп танылғанмен, тілдік ортаның белсенді бірлігіне айналды. Оның орнында «лағып кету» әдеби тіл нұсқасы бола тұрса да, «қуып кеттінің» жаңа ұғымға мағыналануына не түрткі болды деген сұрақтар туындайды.

Керісінше, «ит қуып кетті», «білім қуып кетті» деген бастапқы физикалық қимыл мағынасынан да белсенділігі күштірек болғаны мағыналардың ығысуын байқатады. Қай мағына белсендірек болса, сол қоғамда, сол дәуірде сол мағына үстемдік етіп, бастапқысы ұмытылатын үрдіс қалыптасады. Сондықтан оқу корпусында әрбір сөздің семантизациялану үдерісінің нәтижелері: бастапқы мағыналары, тең мағыналастары, жуық мағыналастары, бейнелі мағыналастары қарама-қарсы мағыналастары, омонимдері, жарыспа тұлғалары бір сызық бойында, яғни қатарлас көрінетіндей құрылуы тиіс.

Демек, оқу корпусына жүктелетін тілтанымдық мәлімет пен оның контентін саралау үшін тіліміздегі сөздердің мағыналану (семантизация) құбылысын зерттеп алуымыз қажет.

Мағыналану (семантизация) – сөздің бірнеше мағыналарға ие болуы, сол мағыналардың бірінің бастапқы, бірінің белсенді, бірінің ауыспалы, бірінің бейнелі, бірінің көне, бірінің жаңа, бірінің ескірген, бірінің ұмытылған күйде ауысуы, яғни өзгеру, толығу, жойылу, т.б. құбылыстары. Сөз мағыналану үдерісіне түскен сайын оның мағыналық қоры кеңейе түседі. Мағыналық қорына орай мағына түрлері де көбейе бастайды. Мағыналану үдерісін қалыптастыратын – тіл тұтынушы, халық. Мағыналану үдерісіне адамның ойлау, тану, сезу, сезіну, қиялдау сияқты когнитивтік әрекеттері себеп болады. Мағыналану үдерісіне синоним, антоним, гипероним, гипоним, энантиосемия, эврисемия, полисемия жатады.

Автоматтандыру – тілтанымдық білімдер жүйесін ғаламтор кеңістігіндегі электронды құралға (қосымша, сайт, іздеу ресурсы, т.б) көпшілікке іздеуге ыңғайлы, тартымды, қысқа әрі нұсқа ақпаратты етіп енгізу.

Ендеше, қазіргі қоғам сұранысына орай тілдік қорды автоматтандырмас бұрын, семантизация құбылысындағы шешілмеген мәселелерді анықтап алу қажет. Себебі тіл үйренушілерге балама парапарлығы аса маңызды. Қазақ тілінің сөздік қоры өзге тілдермен пара-пар келуі тиіс. Соған мүмкіндік жасайтын лексикографиялық қор қолжетімді болу керек. Сондықтан мүмкіндігінше, автоматтандырылған екітілді, тіпті үштілді сөздік қорды әзірлеу қажеттілігі туындайды.

Біріншіден, антонимдер мен қарама-қарсы мағыналы сөздердің параллель лексикографиялық қоры тіл үйренушілер үшін қажет. Бірақ бұл тұста ең алдымен қазақ тілінің антонимдер сөздігі құрастырылуы аса маңызды. 2000 жылы жарық көрген А.Жұмабекованың «Қазақ тілінің антонимдер сөздігі» [170] болғанмен, параллель антонимдер сөздігі құрастырылмаған екен. Бұл тұста мағыналану қағидалары қатаң сақталуы тиіс. Себебі антонимдер лексикалық қабаттан орын алған (жасанды (жұрнақ арқылы) жолмен жасалмаған) градуалды қарама-қайшылықты талап етеді. Мысалы: алыс-жақын, дос-қас, оң-теріс, ауыр-жеңіл, тар-кең, ауру-сау т.б. Егер белгілі бір ұғымды терістеу сипатында қайшылық мағынасын тудыратын болсақ, бұл – қарама-қарсы мағына. Қарама-қарсы мағыналар -сыз, -сіз, -ма,-ме, -ба, -бе, -па, -пе жұрнақтары арқылы туынды сөзден жасалады. Мысалы: ақылды-ақылсыз, көр-көрме, жағымды-жағымсыз. Тіліміздегі антонимдерді осылайша ажырату – мағыналану заңдылығына бағыну деп түсініледі. Антонимдер мен қарама-қарсы мағыналар тізбегі ажырап, толықтырылып болғаннан кейін өзге тілдегі баламасын тіркеу жұмыстары жүргізіледі. Параллель антонимдер сөздігі автоматтандырылып, оқу корпусының бір ұясына салынады.

Екіншіден, синонимдердің автоматтанған нұсқасы – синонимайзер тіл үйренушілер үшін аса маңызды. Синонимдер – адамның сөздік қорын байытады. Қазақ тіліндегі синонимайзердің алғашқы нұсқасы *sozdiqcor* тілтанымдық іздеу веб-сайтына жүктелді. Бірақ оны тіл үйренушілерге тиімді ету үшін параллель (екі не үштілді) нұсқада құрастырған әлдеқайда оңтайлы

болар еді. Екінші жағынан алғашқы синонимайзер – синонимдердің типтік түрлеріне жіктелмеген. Ендігі кезекте синонимайзердің толық нұсқасын әзірлеуде, синонимдік түрлеріне ажыратылған үлгісі басшылыққа алыну керек деп ойлаймыз:

Үшіншіден, фразеологиялық бірліктердің иллюстративті түсіндірмелі параллель сөздігі автоматтандырылса, толыққанды ақпаратты тіл үйренуші бір браузермен тауып алған болар еді. Оқу корпусында фразеологиялық бірліктерді автоматтандыру көрнекілену, эквиваленттелу қағидасы бойынша әзірленеді.

Кесте 15

Автоматтандырылған фразеологиялық қор

Фразеологиялық бірлік	мағынасы	көрнекілігі	Орыс тілінде	Ағылшын тілінде
Түймедейді түйедей ету	Кішкентай мінді, болар-болмас айыпты үлкен етіп, өсіріп көрсетті.		Делать из мухи слона	To make a mountain out of molehill





Төртіншіден, оқу корпусына омонимдер сөздігін автоматтандыру барысында тіл үйренушілерге тиімділігін ескеру қажет. Яғни көрнекілікпен ұсыну жан-жақтылығымен есте тез сақталады.

Кесте 16

Синоним түрлері

Синоним	Лексикалық синоним	Фразеологиялық синоним	Аймақтық синоним	Прагматикалық синоним	Стилистикалық синоним
арам	Жымысқы, харам	Ірі тар	Іші мерез, арамpez, бітінә, жымсыма	жылым құрт	Ішіне пышақ айналмайтын

Омонимайзер үлгісі

ОНИМ	анықтамасы	көрнекілігі
ҚҰР1	Киіз үйдің керегелері мен уықтарын және үзіктің ішкі жағын т. б. байлайтын ені 2-3 елідей, тоқып немесе есіп жасалған жалпақтау бау.	
ҚҰР 2	Таулы және орманды жерлерде жүріп, тіршілік ететін тауық тұқымдас жабайы құс.	
ҚҰР 3	Мінілмеген, семіз, қонды (мал).	
ҚҰР 4	Жылқыны ұстарда шақыратын немесе тоқтарда айтатын ишарат сөз.	

Осы бағытта, қазақ тілінің лексикалық қабатындағы тілтанымдық білімдерді электронды ресурсқа икемдеп форматын түрлендіре беруге әбден болады. Ең басты қағида – тіл үйренушілердің ізденісіне оңай табылатын жағын ескеру маңызды.

Лексикалық бірліктерді автоматтандыру әрине мәтіннің көмегімен жүзеге асады. Сондықтан автоматтандыру үдерісінде мәтін корпусының болуы аса маңызды. Білім беруде лексикалық қабаттың автоматтандырылған қосымшаларын, яғни онлайн-сервистерді пайдалану лексикалық дағдылар мен сөздік қорын дамытуға едәуір септігін тигізеді, бұл дағды тіл үйрену барысындағы тұлғааралық қатынасты жетілдіруге, эмоционалдылыққа, шығармашылық қабілеттерді дамытуға ықпал етеді.

Лексикологиялық білімдерді автоматтандыру үшін ең алдымен мағыналану үдерісіндегі заңдылықтарды тұрақтандыру талап етіледі.

Байқағанымыздай, лексикалық талдау үлгілерін компьютерлік бағдарламалар арқылы автоматтандыратын операциялар жан-жақты тілтанымдық ақпараттарды орындауды қажет етеді. Автоматтандыру нәтижесінде лингвистикалық бағдарламалар қалыптасады. Бүгінгі күнге дейінгі автоматтандырылған тілтанымдық ресурстарды былайша топтастыруға болады:

1. Сөздіктер мен тезаурустар. Олардың қағаз аналогтарының барлық электронды формалары іздеу процесін автоматтандыруға байланысты жұмыс істеуге әлдеқайда ыңғайлы. Сонымен қатар нақты іздеу құралдарымен жабдықталған осындай ресурстардың нұсқаларын құру өзекті болып қала береді.

2. Мәтінді түрлендіру бағдарламалары және мәтін генераторлары. Осы типтегі қарапайым қосымшалар кез келген мәтіндік редакторға енгізіледі. Олар белгілі бір іздеу мүддесін қанағаттандыратын тілтанымдық қормен толықтырылады да, ықтимал қателер жүйесі алдын ала автоматты түрде түзетуге мүмкіндік береді. Сол топқа мәтіндерді автоматты түрде рефераттауға және аннотациялауға арналған бағдарламаларды жатқызуға болады.

3. Мәтіндерді талдау және лингвистикалық өңдеу бағдарламалары. Бұл қосымша емлені тексеруге, трансферттерді орналастыруға, мәтінді лексикалық және стилистикалық тексеруге, олардың қателерінің статистикасын талдауға мүмкіндік береді. Сонымен қатар, күрделі міндеттерінің бірі – мағыналардың өзгеруін, ауысуын, толығын компьютерлендіруге мүмкіндік береді: сөздің лемматизациясы (қалыпқа келтіру, бастапқы формаға келтіру) арқылы көрінеді. Бұнымен қоса тілдің автоматтануының көрінісі – плагиат ресурсы. Лингвистикалық бағдарламалар мәтіндерді автоматты түрде сәйкестендіру және жіктеу үшін утилиталар қосылуы керек. Олар плагиатты анықтауға байланысты мәселелерді шешуге мүмкіндік береді.

4. Табиғи тілді өңдеу жүйелері. Бұл топқа ең күрделі лингвистикалық бағдарламалық жасақтама кіреді. Әрине, ол жоғарыда сипатталған мәселелерді шешуге негізделген, бірақ одан да маңызды мақсаттар қойылады: мәтіннің мағынасын түсіну және сауатты мағыналы жауап беру. Шын мәнінде, бұл жасанды интеллект жүйелерін дамытудың бағыты. Мұндағы негізгі қосымша – компьютерлер үшін табиғи тілдік интерфейс құру.

Қорыта келгенде, оқу корпусына лексикологиялық білімдерді автоматтандыру кезінде ең алдымен мағыналану заңдылығын тұрақты енгізу талап етіледі. Қолданба мүмкіндіктеріне ыңғайлы графикалық интерфейсті қосу талап етіледі. Ақпараттық технологияларға маманданбаған пайдаланушылар үшін интуитивті болуы және тіл үйренушілерге жеңіл болуы талап етіледі.

Қолданбаның жұмыс істеуі әмбебаптығымен ерекшеленуі тиіс. Бір ғана іздеу модулімен шектелмей, ықтимал нұсқалары сілтемеленуі маңызды. Мысалы: қазақ тілінде емле ережесі бойынша немқұрайлы ма немқұрайды ма? Оны тауып беретін қолданба қосымшада жарыспа тұлғасы әдеби нұсқасына сілтенуі қажет [171]. Бұл өңдеу тапсырмалары қатарына енеді. Демек, мысалдарда бағдарламаның барлық модульдерінің жұмыс нәтижелері көрсетіледі.

7 ДИАЛЕКТІЛІК МӘТІНДЕР ІШКОРПУСЫ

7.1 Қазақ диалектологиясы бойынша зерттеулердегі негізгі бағыттар

Соңғы жылдар ішінде қазақ тіл білімінің диалектология саласы лексикография (диалектологиялық, аймақтық сөздіктер түзу), лингвистикалық география, этнолингвистика, диаспоратанумен (диаспорология) байланысты бағыттарда дамып келеді.

Өткен ғасырдың 90 жылдарының ортасында белгілі диалектолог ғалымдар Ш.Сарыбаев, Ә.Нұрмағанбетов, С.Омарбеков, О.Нақысбеков, Ғ.Қалиев т.б. қатысуымен үш томдық етіп жоспарланған академиялық «Қазақ тілінің диалектологиялық сөздігінің» бірінші және екінші кітаптары жарыққа шықты [172]. Сөздік Қазақстанның әр облыстарындағы диалектілік сипаттағы лексикалық, фразеологиялық, фонетикалық ерекшеліктерді қамтитын, осыған дейін 1969 жылы шыққан «Диалектологиялық сөздіктің» толықтырылған нұсқасы болатын. Бұл еңбек бұрын-соңды зерттелмеген Қазақстанның орталық, солтүстік облыстары мен Тәжікстан, Моңғол Халық республикаларынан жиналған тілдік деректерді қамтыды.

Мемлекеттік «Мәдени мұра» бағдарламасы қолға алынған тұста жоғарыда аталған сөздіктердің негізінде және соңғы 10-15 жылдарда жүргізілген диалектологиялық зерттеулер, жарияланған мақалалардағы материалдардың басын қосып, толықтырудың нәтижесінде 2005 жылы «Қазақ тілінің аймақтық сөздігі» жарық көрді [173]. Сөздіктің негізгі материалы ретінде Қазақстан территориясынан тысқары, шет елдерде өзге тілдік ортада шоғырланып орналасқан қазақтардың, нақты айтқанда, Өзбекстан, Қарақалпақстан, Түрікменстан, Тәжікстан, Ресей федерациясындағы, сонымен қатар алыс шет елдер Моңғолия, Қытай, Ауғанстан, Иран елдеріндегі қазақ диаспорасының тілінен лексикалық, фонетикалық және грамматикалық та ерекшеліктер пайдаланылды. Шетел қазақтарының лексикасын зерттеу, оның диаспоралық, аймақтық сөздігін жасау мемлекеттік тілдің дамуына, оның нормаларын тұрақтандыруға зор үлес қосады. Диалектология саласындағы лексикографиялық үрдіс «Тілдерді қолдану мен дамытудың 2001–2010 жылдарға арналған мемлекеттік бағдарламасы» аясында жалғасып, аймақтық сөздіктің ықшамдалған, бейімделген түрі – «Диалектологиялық сөздік» те көпшіліктің назарына ұсынылды. Бұл сөздікте негізінен тіліміздің лексикалық әлеуетін көрсететін, әдеби тілдің баю көздері болып табылатын сөздерге, тарихи-танымдық мәнділікке ие атауларға, аймақтық сипатта болғанымен, әдеби тіл құрамына ене қоймаған түрлі шаруашылық салаларында қолданылып жүрген кәсіби сөздерге баса мән берілді.

Тіл білімінің диалектология тармағының дербес саласы ретінде лингвогеография XIX ғасырдың соңғы ширегінде қалыптасты. Әр алуан тілдердегі диалектологиялық құбылыстардың көптеп жиналуы мұндай құбылыс атаулының тараған шегін анықтауды тіл білімпаздары алдына міндет етіп қойды. Зерттеудің бұл әдісінің мәні – нақтылы тілдік құбылыстарды географиялық

картаның бетіне түсіріп, сол арқылы оның тараған жерлерін анықтау. Тіл құбылыстарының тараған аймағы мен шегін айқындау тілдік жүйенің пайда болу, даму заңдылықтарын танып-білуге, олардың нақтылы географиялық кеңістікке орналасу себептері мен сырын түсінуге, түптеп келгенде, тілдің тарихын түсінуге мүмкіндік береді.

Қазақ лингвистикалық географиясы қазіргі А.Байтұрсынұлы атындағы Тіл білімі институтында 60-жылдардың орта шенінен бастап жоспарлы түрде қолға алынды. Жинақталған деректер негізінде бір топ қазақ диалектологтері 70-жылдардың бас кезінде аталған өңірлерден жиналған тілдік құбылыстардың арасынан басты-басты дегендерін іріктеп алып, карта бетіне түсіріп, 250-ден астам лингвистикалық карта жасап шығарды. Қазақ тілінің диалектологиялық атласын жасаудың бірқатар методикалық мәселелерін сөз ететін Ш.Сарыбаевтың алғашқы мақаласы, Т.Айдаровтың «Лингвистикалық география» деген, сосын Өзбекстан жеріндегі қазақтардың тілін лингвистикалық география тұрғысынан зерттеуге арналған кітаптары, атласқа деп жиналған фонетикалық ерекшеліктерді ішінара түсіндіруге арналған О.Нақысбековтің [174] мақаласы да «Қазақ тілінің кіші лингвистикалық атласын» жасауға негіз болды.

Қазақ тілінің лингвистикалық атласының картографиялық жұмысын жүргізуде мынадай деректердің болғандығын айтуымыз керек:

– тікелей лингвогеографиялық мақсатпен 60-жылдарда оңтүстік-шығыс өңірден жиналған диалектологиялық деректер;

– осы деректер негізіндегі диалектологиялық карталардың қолдан жазылған алғашқы нұсқасы;

– 70 жылдарда бұрынғы «Кеңес одағындағы түркі тілдердің лингвистикалық атласы» үшін Қазақстан территориясынан енгізілген диалектологиялық деректер мен фонетикалық құбылыстарға арналып баспа жүзінде жарияланған лингвистикалық карталары;

– А.Байтұрсынұлы атындағы Тіл білімі институтының бұрынғы Қазақ диалектологиясы бөлімінің картотека қорында жинақталған деректер;

– диалектологиялық сөздіктердің дерегі;

– қазақ диалектологиясы бойынша монографиялар мен мақала зерттеулердің дерегі;

– тікелей қазақ лингвогеографиясына арналған монографиялар мен мақала зерттеулердің дерегі;

– атласты құрастырушылардың қырық жылдан астам уақыт қазақ диалектологиясымен айналыса жүріп, өз жинап-тергендері мен іс тәжірибесі.

Осы деректерді кеңінен пайдалана отырып белгілі диалектолог ғалымдар С.Омарбеков мен О.Нақысбеков қазақ тілінің фонетикалық, грамматикалық, лексикалық жүйесін лингвогеографиялық тұрғыдан сипаттауға арналған карталар сериясын құрастырып шығарды [175, 220]. Бұл еңбекке осы карталардың ішінен іріктелініп алынған 300 карта енді.

Қазақ тіліндегі диалектілік ерекшеліктерді зерттеу жұмысы соңғы жылдары Республика көлемінен шығып, алыс-жақын шетелдерде тұратын қазақ диаспорасы тілін де қамти бастады. Диалектологиялық еңбектерде шетелде-

гі қазақтардың тілі қазақ тілінің белгілі бір сөйленістер (говор) тобына жататын вариант (сөйленіс) ретінде сипатталып, тілдік ерекшеліктері жинастырылып, оларға талдау жасау арқылы қолданылу жиілігі, өзге сөйленістерімен ұқсастықтары мен айырмашылықтарын салыстыру тұрғысынан зерттелді. Сонымен байланысты кейінгі 10-15 жылдың ішінде талай жыл алыс шетелдерде, өзге тілдік ортада өмір сүрген қазақ тілінің жай күйін, тілдегі өзгерістерді зерттеу мүмкіндігі туды. Әртүрлі географиялық әрі тілдік ортада өмір сүріп отырған қазақтар тілінің ерекшелігі әрқилы. Ол ондағы аралас отырған халықтардың, үстем тілдің әсерімен байланысты.

Отанынан жырақта өзге мемлекеттерде қоныстанған белгілі бір ұлттың шоғырының орналасу тарихын, жай-күйін әлеуметтану ғылымының диаспоратану саласы зерттейді. Диаспоратануда шетелдегі қазақтарды қоныстану ерекшелігіне байланысты қазақ диаспорасы мен ирридені деп қарастырылады. Зерттеулерде бұрынғы Кеңестер одағының құрамында болған республикалар, бүгінде Ресей, Өзбекстан, Түрікменстан, Қарақалпақстан республикаларымен бірге алыс шет елдер Моңғолия, Қытай, Ауғанстан, Түркия т.б. елдердегі қазақ диаспорасы мен ирриденінің тілін зерттеу алға қойылды. Бұл елдермен қатар қазақтар Еуропа, Пәкістан, Ливия т.б. елдерде бірен-саран қоныстанғанымен негізінен зерттеу нысанына Моңғолия, Қытай, Ауғанстан, Иран елдеріндегі, соңғы кезде Түркиядағы қазақтардың тілі алынды, тіл ерекшеліктері бойынша зерттеу жұмыстары қорғалып, мақалалар, монографиялық зерттеулер жарық көрді. Осылардың ішіндегі кейбір зерттеулерге тоқталып өтсек.

Шығыс өңірі қазақтарының тілдік ерекшелігін зерттеген Ж.Болатов Ресейдің Таулы Алтай өлкесіндегі қазақтардың тілін Қошағаш, Қытай қазақтарының тілін Шыңжаң, ал Моңғолия қазақтарының тілін Баян-Өлгий сөйленістері деп, оларды Ертіс өзенінің екі жақ бетінде орналасқан Өскемен және Аягөз сөйленістерімен қосып шығыс сөйленістер тобы деп қарастырды. Моңғолия қазақтарының тілін арнайы зерттеу нысаны етіп, жан-жақты зерттеген ғалым Б.Базылхан еңбегінде [176] тұрғылықты қазақтардың тілінде жергілікті ерекшеліктермен қатар қазіргі халықтық тілімізде ескіріп, қолданыстан шыққан сол кездердегі байырғы сөздердің біразы әлі де қолданылады. Оның пікірінше мұндағы қазақтар тілінің сақталуының басты бір себебі олардың Моңғолияға қоныс аудару процесінің әдеби тіліміз әбден кемелденген тұсқа тура келуімен байланысты. Кейінгі жылдары Моңғолиядағы қазақтар тілінің дыбыс жүйесі, ондағы тілге түрлі сыртқы фактордың әсерін, қалыптасқан тілдік жағдаятты, көстілділікті әлеуметтік лингвистика тұрғысынан зерттеу жүргізілді [177].

Алыс-жақын шет елдерде тұратын қазақ диаспорасының тілі дегенде, біздің ең басты арқа сүйейтініміз, әрине, Шыңжаң өлкесіндегі қазақтардың тілі. Бұлай деудегі негіз, біріншіден, Шыңжаң өлкесінде тұрып жатқан қазақтардың саны біршама көп (кейбір болжалдар бойынша екі миллионға жуық); екіншіден, Шыңжаңдағы қазақтар тілінің еркін даму мүмкіндіктері болды. Бұл өлкеде қазірге дейін қазақ тіліндегі мектептер ашылды, газет, журналдар, кітап, оқулықтар баспадан шықты, радио, телевидение жұмыс істеп келді. Сонымен бірге бұл өлкедегі қазақ тілі біршама жүйелі зерттелді. Шыңжаңдағы қазақтардың тілі диалектілік құбылыстардың молдығы әрі күрделілігімен ерекшеленеді.

Шыңжаңдағы қазақтар тілін зерттеу жалпыхалықтық қазақ әдеби тілін мыңдаған жеке сөздер, фразеологизмдер, мақал-мәтелдер, тәмсілдер және әр саладағы ғылыми терминдермен өзгешеленетіні даусыз. Біздіңше, Шыңжаңдағы қазақтардың тіліне екі түрлі фактор әсер еткендігі байқалады. Оның бірі, сыртқы фактор, яғни ғасырлар бойы қоян-қолтық аралас өмір сүріп отырған қытай, ұйғыр, азды-көпті моңғол тілдерінің әсері болса; екіншісі, жалпыхалықтық тілді негіз ете отырып, тілдің ішкі даму заңдылығы бойынша дамуы. Қазірге дейін Шыңжаңдағы қазақ тілінің ішкі даму заңдылығы бойынша дамуы басым орынды болып келді [178, 78].

Шыңжаңдағы қазақтар тілін жалпыхалықтық қазақ тілімен салыстыра зерттегенде, олардың арасындағы фонетикалық, лексикалық азды-көпті грамматикалық өзгешеліктерге тоқталған Б.С.Нұрғабыл, Қ.Ғабитханұлы, С.Мұстафаұлы т.б. зерттеу мақалалары, монографияларын атауға болады.

Келесі бір зерттеуге негіз болып отырған нысан қиыншылық кездерде Ауған мен Иранға ауа көшіп, сол жерді сексен жылға жуық мекен етіп, тұрған Отаны Қазақстанға қайта көшіп келе бастаған қазақтар тілінің ерекшелігі. Ауғанстан мен Иран қазақтарының тарихын, этнографиясын, этникалық тобы мен тілдік ерекшелігін жинау 1990 жылдан кейін ғана басталды. Ауғанстан қазақтары пуштулармен, тәжіктермен, өзбектермен, түрікмендермен бірге өмір сүріп келген болса, Иран қазақтары парсылармен, түрікмендермен тығыз қарым-қатынаста болған. Соның нәтижесінде бұл өлке қазақтары біраз тілдік, этнографиялық ерекшеліктерге ие болған. Ауғанстан мен Иранда тұратын қазақтар тілін өз алдына дербес биллингвизм сипатындағы ерекшелік деп қарастыруға мүмкіндік береді. Ауған қазақтары пушту, парсы, өзбек, түркімен тілдерін, Иран қазақтары парсы, түркімен тілдерін ана тіліндей меңгерген қазақ халқының негізгі аудандарынан жырақ жатқан, оның мәдени-экономикалық өмірі мен байланысы жоқ басқа тілдік говорларда қазақ әдеби тілінің әсері солғын болатынын қазақ сөйленістерін зерттеген көптеген лингвистер дәлелдеді [179, 15].

Дегенмен алыс шет елдердегі қазақ тілінің материалдары, түрлі көркем әдебиеттердегі, жырлардағы, фольклордағы мәтіндер, фразеологизмдер, пареомологизмдер, шұғылданатын кәсібіне қатысты, жергілікті тілмен қарым-қатынаста пайда болған аймақтық бірліктер жан-жақты, толық жиналып бітті деуге болмайды. Ендігі кезекте мұндай зерттеуді жүргізу үшін Қытайдағы, Түркиядағы, Моңғолиядағы, Ауғанстан, Иран елдеріндегі қазақ шоғырының тілін, ондағы тілдік жағдаятты әлеуметтік-диалектологиялық аспектіде зерттеу өз кезегін күтуде. Себебі уақыт озған сайын түрлі саяси-әлеуметтік факторларға, тілдік жағдаятқа, қоршаған тілдік ортаның әсеріне байланысты сақталынып келген қазақ тілі де (Қытайдағы, Моңғолиядағы) өзгеріске ұшырауда. Ал кейбір елдегі (мысалы, Түркияда) қазақтар тілінің жай-күйі тіпті белгісіз жағдайда, яғни әлі зерттелген емес.

Аталған елдердегі қазақ тілінің жай-күйін, дамуын, өзгерісін, қолданысын зерттеп көрсету үшін ондағы қазақтардың тілі тек қана сөйленіс ретінде емес, қалыптасқан термин, атаулары бар, өзіндік әдеби нормаға түскен, жазбаша, ауызша түрде қолданылатындығын ескере келе ондағы қазақ тілінің тілдік

мәртебесі анықталуы тиіс. Өйткені Қытай, Моңғолия, Түркия елдерінде қазақ тілінде мерзімді басылымдар, саяси-көркем әдебиеттер, оқулықтар шығады, радио, телехабарлар таратылады. Ауызекі тілдегі айырмашылықты айтпағанда, шет елдерде қазақ тілінде жарық көрген әдебиеттердің Қазақстанда қолданылып отырған әдеби тіліміздің жазу нормасымен сәйкес келмейтін тұстары жеткілікті. Сонымен қатар аймақтардағы басқару, техника, шаруашылық, ғылым, саясат т.б. салаларда жұмсалатын атау-терминдерде бірізділік жоқ. Әр аймақта тілдердің өзара әсерлесу арқылы кірме лексикалық бірліктер, өзіндік терминдер қолданылады. Қытайдағы, Моңғолиядағы, Түркия қазақтарының тілінде кірме элементтермен қатар сәтті қолданысқа енген төл терминдер де бар. Мұндай фактілерді тіркеп, жинастыру арқылы терминологиядағы ортақтықты айқындау, сәтті қолданыстарды Қазақстандағы әдеби тілдің игілігіне айналдыру да кезек күттірмейтін мәселе.

Аталған елдердегі қолданылып отырған қазақтардың жазу-графикасы да бірдей емес. Бір елде (Моңғолия) кириллица, енді бірінде (Түркия) латын, бірінде (Қытай) төте жазу қолданылып келеді. Алдағы уақытта бұл мәселе де зерттеу нысаны болуы тиіс. Әрі жазуымыз латын графикасына көшкен жағдайда бұл елдердегі жазу проблемасын, тәжірибесін кеңінен зерттеудің септігі тиетіні анық.

7.2 Диалектілік мәтіндер ішкорпусы: әзірлемесі мен технологиясы

Диалектілік корпус тілдегі жергілікті ерекшеліктерді тілдік-құрылымдық және сөйлеу/дискурстық сипатта зерттеуге қызмет ету керек. Осы ретте басқа елдердің диалектілік корпустарында әдетте диалектілік ішкорпустар жергілікті тілдегі сөйлесімдерді (аудио, видеожазбаларды) жазып алып, аудио және олардың стенограммасын (мәтінін) корпус базасына салады. Өйткені жергілікті ерекшеліктер лексика (сөз) деңгейінде ғана емес, жергілікті дискурстан да байқалады. Лексикалық ерекшеліктер жергілікті дискурсты құраушы сөздік қор болса, қарым-қатынас кезінде бұл лексикалық қордың грамматикалық құбылуы мен контекстік қолданысы, просодикалық айтылымы, прагматикалық мәні сол жергілікті халық тілінде ерекшеленіп отырады. Сондықтан ұлттық тілдегі жергілікті ерекшеліктер тілдік бірліктердің динамикасында айқындала түседі. Сонымен қатар табиғи тілдегі жергілікті тілдік ерекшеліктер тіркелген сөздерден, яғни сөздіктерде берілген диалектизмдерден ғана емес, күнделікті қарым-қатынас кезіндегі сөйлесімдерден анық көрінеді. Осы ретте диалектілік корпусстарды тілдерде жергілікті дискурстан жазып алып, оларды корпус базасына жинақтап жасаған. Мұндай жергілікті сөйлесімдер сол халық тіліндегі сөйлеу ерекшеліктерін ғана көрсетпейді, сонымен қатар мәдениетін, дүниетанымын, менталитетін де байқатады. Өйткені тіл – ұлттың, жергілікті елдің дүниетанымын көрсететін, репрезенттейтін бірден-бір құрал болып табылады.

Жергілікті тілдегі сөйлесімдерді жазып алу сөйлесімнің әртүрлі жағдайларынан, әртүрлі ортадан, әртүрлі сөйлеу түрінен алынады. Тілдік қарым-

катынас жасаудың қанша типі болса, солардың барлығы да жергілікті елдің тілдік ерекшелігінен, дүниетанымынан, мәдениетінен ақпарат береді. Орыс тіл білімінде диалектілік сөйлеу тілін зерттеу XX ғасырдың аяғы мен XXI ғасырдың басында белсенді жүргізіле бастады. Бұл ретте Качинская И.Б. [180, 57 б.], [181, 245 б.], Летучий А.Б. [182, 215 б.], [183, 114 б.], Сичинава Д.В., [184, 58 б.] т.б. ғалымдар еңбектерін атауға болады. Ғалымдар диалектілік сөйлеу тілін зерттеудің маңызы туралы былай деп жазады: «Диалектная речь (рассказы диалектоносителей, обиходно-бытовое общение на диалекте, диалоги сельских жителей с представителями иной социокультурной среды) – важнейший транслятор традиционной народной культуры. Если такие лингвокультурные феномены, как устное народное творчество, системы обрядов и ритуалов, представляют культурную традицию скорее в ее диахронном срезе, сохраняя и транслируя в символически обобщенном и эстетически обработанном виде сложившиеся на протяжении длительного времени элементы мировидения и миропонимания, то актуальная этнокультурная информация (современная картина мира носителей традиционной народной культуры) репрезентирована прежде всего современной диалектной речью, в которой в виде актуализованного знания могут также объективироваться и элементы других семиотических систем, параллельно существующих в данном социуме и в сознании его представителей. В связи с этим лингвокультурологическое изучение народных говоров является актуальной задачей современной диалектологии» [185].

Демек, ұлт тіліндегі жергілікті ерекшеліктер сөйлеу тілінен жазылып алынғанда сөйлесімге түсушілердің прагматикасын және дүниетанымы мен менталитетін байқауға болады. Осы тұрғыдан алғанда жергілікті ерекшеліктерді тілде тіркеп қана қоймай (яғни лексикасын сөздіктерде белгілеу), аудиотаспаға түсіріп, корпус базасына жинақтау қажет. Бұл табиғи тілдегі жергілікті сөйлеу ерекшеліктерін және өңір тұрғындарының бітім-болмысын терең тануға мүмкіндік береді. Диалектілік-репрезентативті корпус – жергілікті сөйлеу тілін құрылымдық, коммуниктивтік, когнитивтік, лингвомәдениеттанымдық тұрғыдан зерттеудің көзі. Мәселен, Н.Г.Чернышевский атындағы Саратов мемлекеттік университеті жасаған Саратов диалектілік корпусы осы жоғарыда аталған зерттеу дереккөзі ретінде арнайы жасақталған. Бұл корпуста жергілікті тілдегі ауылдық сөйлесім жазбалары жинақталған. Бұл диалектілік корпус үш ішкорпуста тұрады. Біріншісі – Саратов облысы, Вольск ауданы, Белогорное селосының говоруы, екіншісі – Аткар ауданы, Земляной Хутор ауылының говоруы, үшіншісі – Вологод облысы, Вытегор ауданы, Мегра селосының говоруы. Бұл диалекторлогиялық корпуста жергілікті аудиожазбадағы сөйлесімдермен бірге экстралингвистикалық деректер де салынған. Мысалы, олар видеоиллюстрациялар, карта, сызба, фотосуреттер, географиялық мәліметтер, өңір туралы тарихи мәліметтер т.б. Олар корпуста ақпараттық тізбек ретінде жалғаспалы түрде орналастырылған.

Саратов диалектілік корпусында говорлардан жазылып алынған аудиожазбалар үш түрде сақталады: 1 – аудиожазба, 2 – аудиожазбаның символдық жазбасы/мәтіні, 3 – белгіленім қойылған мәтіндік модуль ретінде.

Корпус әзірлемешілерінің айтуынша, жазылған аудиолар мәтіні орфографиялық нормаға жақын жазылған, бірақ жергілікті тілдегі лексикалық, грамматикалық диалектілік ерекшеліктер сақталып жазылған. Мысалы:

Вот это у нас образа/ оне называются старообрядческие/ поморские/ вот наша вера какая/ мы не монашки/ мы не какие поповцы/ и не это... вот оне эти вот/ монашки-ти/ оне... как сказать... оне... тоже беспоповцы/ но... у них брак/ и считается за грех/ оне называются безбрачные// оне вот доживают/ до шестьдесят лет/ женщина/ если с мужчиной живет/ это было раньше/ и у них закон такой/ доживают до шестьдесят лет/ она сносит кануны/ и говорит/ вот Ваня там или Вася/ всё/ я больше прекращаю с тобой жизнью жить/

Корпустың метабелгіленімінде аудиожазбаның жанры мен тақырыбы көрсетілген. Мысалы, аудиожазба әңгіме, ертегі, мақал-мәтел, ән т.б. болуы мүмкін, ал тақырыбы – отбасы, дін, табиғат, оқиға т.б. Корпус әзірлемешілері аталған жанр мен тақырыптық топты белгілеуде олардың түрлерін санмен кодтап барып (#25@1) қойған. Берілген материалдарынан бұл метамәтіндік ақпараттарды лингвистикалық (морфологиялық) белгіленім қатарында қатар береді. Мысалы:

#25@1ну {ну(ну)=PART} <...> когда {когда(когда)=ADV/ CONJ} стало {стать(стать)=V,сов=изъяв,прош,ед,сред} мне- {я(я)=S-PRO,ед,од=дат} годов {год(год)=S,муж,неод=мн,ро д=*} шестнац ать {шестнадцать(шестнадцать)=NUM=им} или {или(или)=CONJ} пятнадцать {пятнадцать(пятнадцать)=NUM=им}/ это {это(это)=PART} уже {уже(уже)=ADV} пошла {пойти(пойти)= V,сов=изъяв,прош,ед,жен} мода {мода(м ода)=S,жен,неод=ед,им} панталоны {панталоны(панталоны)=S,жен,неод,мн=вин} белые {белый(белый)=A=мн,вин,нео д} с {с(с)=PR} кружевами {кружева=S,сред,неод,мн=твор} // <...> и {и(и)=CONJ} нижни е {нижний(нижний)=A=мн,вин,н еод} юбки {юбка(юбка)=S,же н,неод=мн,вин} белые {белый(б елый)=A=мн,вин,неод} // <...> чтобы {чтобы(чтобы)=CONJ} немного {немного(немного)=ADV}/ из-под {из-под(из-под)=PR} п латьев {платье(платье)=S,сре д,неод=мн,род} <...> видать {ви-дать(видно)=PRAEDIC=*} было {быть(быть)=V,несов=изъяв ,прош,ед,сред} // и {и(и)=CONJ} начали {начать(начать)=V,сов =изъяв,прош,мн} уже {уже(уже)=ADV} юбки {юбка(юбка)=S, жен,неод=мн,вин} – то {то(то)=PART} короткие {короткий(кор откий)=A=мн,вин,неод} носить {носить(носить)=V,несов=ин ф}/ а+то {а то(а то)=CONJ} раньше {раньше(раньше)=ADV} –то {то(то)=PART} не {не(не)=PART} были {быть(быть)=V,несов= изъяв,прош,мн=*} // @1#25

Тақырыптық, жанрлық кодтардың аудиожазба мәтінінде берілуі сол белгіленім бойынша пайдаланушыға іздеу жасауына мүмкіндік береді.

Әрбір мәтінге метамәтіндік сипаттағы 3 модуль тіркеліп отырады. Олар:

1) Мәтін метабелгіленімі (информанттар туралы мәліметтер, мәтін жазылып алынған уақыт, мәтін жазылып алынған жер, қатынас жағдаяты, сөйлесім адресаттары, сөйлесім тақырыбы (кенестік дәуірге дейін, революция, Азамат

соғысы, коллективизация, Ұлы Отан соғысы, соғыстан кейінгі кеңестік дәуір, кеңестік дәуірден кейінгі кезең; 2) информант биографиясы; 3) иллюстративтік модуль (информант фотосы, алынған мәтінге қатысты фотоиллюстрациялар).

Саратов диалектілік корпусында осы аталған метамәтіндік мәліметтер мен лексика-морфологиялық белгіленімдер іздеу жүйесіне салынып, біртұтас программалық қызмет түрінде жұмыс істейді. Корпустың мәтіндік базасында алынған мәтіндер жазылып алынған жердің жергілікті халқының өмірінен, тарихынан, менталитеті мен өмірлік көзқарасынан хабар береді [186].

Саратов корпусында жазылып алынған аудиожазбалар тақырыптық тұрғыдан көбінесе тұрмыстық өмір, үй, үй шаруашылығы, дін тақырыбындағы ірі макротоптар болып келген. Саратов корпусы материалдарын диалектологияны, дискурс, ауызша тіл, стилистика, тілдік қатынас, когнитология, лингвомәдениеттану мәселелерін зерттеушілер өз мақсаттарына сәйкес тілдік дереккөз ретінде пайдаланған. Басқа да елдердің диалектілік ішкорпустарын ғалымдар әр қырынан дереккөз етіп алып, түрлі зерттеулер жүргізуде. Мысалы, диалектілік корпус базасын қолдану арқылы зерттеушілер қала өмірі мен ауылдық жердегі психологиялық, дүниетнымдық айырмашылықтарды анықтай алған, ауылдың өмір сүру дағдысын, түсінігін, өмірлік тіршілігін, сана-сезімін, сөйлеу ерекшелігін т.б. толып жатқан когнитивтік, мәдени, діни, ділдік, тілдік ерекшеліктерді айқындаған. Ауыл-қала оппозициясының моделін жасаған.

Томск диалектілік корпусы лексикалық және мәнмәтіндік тұрғыда қарастырылған. Мәтінге негізделген Томск диалектілік корпусы белгіленімнің жаңа түрін қажет етті, диалектілік дискурста мәтіннің ұйымдасу сипаттамасын көрсетеді. Ол шартты түрде мәтін типі бойынша белгіленім деп аталды.

Диалектілік дискурстың ерекшеліктері әлі де толық зерттелмегендіктен, Томск диалектілік корпусы бұл кезеңде ауызша сөйлеудегі тұрақты маркерлердің негізінде жүзеге асырды. Осындай әртүрлі белгіленімдерге мыналар енгізілді:

а) Сөйлеу құбылыстарының машықсыз айтылуы деңгейі бойынша мәтін түрлері.

б) Сөйлеу жанрлары.

Бірінші белгіленімге мәтін түрінде диалектілерді қолданушылар арасындағы халықтық сөйлеу мәдениеті өкілдерінің диалектілерді сөйлеу формасына тән мәжбүрлеусіз қолдануындағы қарым-қатынасы негізінде қарастырылды. Яғни диалектологтар арасында белгілі бір тақырып бойынша мақсатты түрде жүргізілетін талқылаулар, әңгімелесу барысындағы вербалды қимылдар (*Ой, немерем оянды*) әңгімені таратып айтуы кезіндегі ақпарат берушілердің тақырыбының өзгеруі кезіндегі қосынды сөздер (Сізге не айтсам екен? Дәм алыңыз, ауыз тиіңіз). Алдыңғы тип текстерінен өзгешелігі жалпы тақырыпқа бағытталғандығы («Өсімдіктер атауы», «Салттар» т.б.). Ал екінші сөйлеу жанрлары бойынша Томск диалектілік корпусында материалдардың толық белгіленімі ниет сипаттамасы (ақпараттық, императивті, рәсім және бағалау) бойынша жанр типологиясына сәйкес беруді көздеді; әр жағдаят бойынша нақты жанр көрсетілді (ақпараттық топқа әртүрлі хабарлама түрлері кірді, болжам, түсініктеме, арыз, ескерту; императивті – өтініш, тапсырыс, тапсыр-

ма, бұйрық, алғыс, шақыру, тілек; бағалау – мақтау, айыптау, өзін-өзі бағалау, бағалау). Сонымен қатар биографиялық әңгімелер сюжетті әңгіме, суреттеу, сұхбат, ертегі т.б. қамтылды [187, 58 б.].

Алайда бұл жанрлардың бәрін қамту мүмкін болмағандықтан, барынша жиі кездесетін жанрларды көрсету керектігі қарастырылды. Яғни олар автобиографиялық әңгіме – ақпарат берушінің ресми емес өмірі туралы өздерінің айтқан әңгімелері, еске алу – адамның өткен өмірі туралы монологта айтылатын оқиғаларды баяндаудағы сөйлеу жанры, басқа адамдар туралы айтылған әңгімелер; ақпарат берушінің өзі немесе жақын туысының басынан өткізген өмірлік оқиғасы туралы әңгімелерге белгіленім қойылды. Халықтың сөйлеу мәдениетінде ерекше орын алатын ертегі, мақал-мәтелдер т.б. жанрларына да назар аударылған.

Томск диалектілік корпусын құрастырушылар Орта Прибыя говорларын қолданушылардың лексиконындағы бай сөздік қорын, жарық көрген диалектілік сөздіктерді цифрландыру бойынша жұмыстар жүргізуде. Томск диалектілік корпусы бірінші қадам жолында лексикалық белгіленімге әдеби емес бірліктер мағынасының түсіндірмесін беруді қамтамасыз етуді жоспарлап отыр. Болашақта «корпус - сөздіктер» байланысын кеңейте отырып, басқа да сөздік көздерін пайдаланып, лексемалардың сипаттамасын (көнерген, жаңа, балалық, дөрекі т.б.) аша түсуді көздеп отыр. Бұл мәліметтер жаңа корпусты қолдана отырып, диалектілік лексиканы зерттеуде және жаңа лексикографиялық еңбектер құрастыруда қолданылатын тиімді құрал болып табылады. Томск диалектілік корпусын жасау өз кезегінде осы уақытқа дейінгі диалектілік материалдар мен сөздіктердің фундаменти болмақ.

Орыс тілінің ұлттық корпусында Диалект мәтіндер ішкорпусы 2005 жылдан бастап сайтқа салынған. Олар да диалектілік сөйлеу тілін аудиотаспаға жазып, орфографиялық стенограммасын жасаған. Ғалымдардың айтуынша, бастапқы кезде жазылып алынған мәтіндердің метаақпараты (жазылып алынған уақыт, жері, реципиенттер, тақырыбы т.б.) берілген де, ондағы сөздерге лингвистикалық белгіленімдер қойылмаған. Кейін 2009-11 жылдары әрбір лексеманы лингвистикалық өңдеуден өткізу қолға алынған. Орыс тілінің бұл диалектілік корпусына Ресей өңірінің әр жерінен таспаға жазылып алынған аудиожазбалар енгізілген. Тіпті осы ішкорпус материалдары бойынша ғылыми-зерттеу жұмыстары да жүргізілуде [188, 14 б.].

Орыс тілінің диалектілік корпусында аудиотаспаға жазылып алынған мәтіндердің түпнұсқа аудиосы мен екпін сақталған фонетикалық транскрипциясы да енгізілген. Барлық диалектілік мәтіндерге лингвистикалық және метамәтіндік (жанр, тақырып, фонетикалық ерекшелігі (жалпы түрдегі) белгіленімдер жасалған.

Орыс тілінің диалектілік корпусында метамәтіндік белгілінем бойынша іздеу мына параметрлерді қамтиды:

- 1) Жазба географиясы (өлке: облыс, аудан)
- 2) Тақырыбы
- 3) Жанры
- 4) Жазылған уақыты

- 5) Тарихи кезеңі, суреттелген оқиғаның кезеңі
 - 6) Кейбір фонетикалық ерекшеліктер
 - 7) Информанттар туралы ақпараттар (жынысы, жасы, білімі, конфессиясы)
- Лингвистикалық белгіленім бойынша:

- 1) Грамматикасы (диалектілік ерекшеліктер қамтылған)
- 2) Семантикасы (арагидік)
- 3) Лексема бойынша іздеу

Орыс тілінің диалектілік корпусында грамматикалық белгіленімдер қоюда 2 түрлі тәсіл қолданылғандығы айтылады. Біріншісі – орыс тілінің говорларына тән грамматикалық қосымшаларындағы ерекшеліктерді әдеби тілдегі орфографиялық нормаға сәйкес инвариантын тауып, негізгі корпустағы морфологиялық анализатор арқылы талдау. Мұндайда орфографиялық редактор (подстрочник) транскрипцияны әдеби нормаға келтіреді. Орфографиялық редактор транскрипцияны орфографиялық нормаға келтірген соң, негізгі корпуста қолданылатын морфологиялық анализатордан өткізіледі. Анализатор грамматикалық белгіленімдер қоюда таба алмаған жағдайда оны диалектілік грамматикалық ерекшелік ретінде көрсетеді.

Сонымен қатар негізгі реестр сөздігінде жоқ сөздер диалектілік мәтіндерде табылмаған сөз болып айырықшаланады.

Екіншісі – говорларға тән грамматикалық түрленім сөздігін корпус базасына салу арқылы морфологиялық анализатордан өткізу. Орыс тілінде мұндай грамматикалық ерекшеліктер септікте, видта, көптікте, тек категориясында т.б. кездеседі екен. Мысалы, императив формасында постави (поставь), удари (ударь), посоль (посоли). Мұны диалектілік форма деп атайды. Ал егер жергілікті ерекшелік лексема/ негіз деңгейінде болса, диалстем деп атайды. Орыс тілі говорларына диалектілік флексия да тән.

Орыс тілінің диалектілік корпусын әзірлеушілері айтуынша, бұл корпусқа тек өңірлерден жазылған ауызша мәтіндер ғана кірмеген. Өңірлік газеттер, өңірлік хаттар, күнделіктер, өлеңдер, мемуарлар, жазба кітапшалар (записные книжки) да кірген.

Орыс тілінің диалектілік корпусына жинақталған мәтіндерді төмендегіше бөліп көрсетеді:

1) Фольклорлық емес ауызша мәтіндер (диалогтар, әңгімелер, баяндаулар (пересказ), сұраққа жауаптар, пікірталастар т.б.);

2) Фольклорлық емес жазба мәтіндер: автобиография, еске алу/мемуарлар, жеке хаттар, құттықтаулар, жеке күнделік, жазба кітапшалар, қызметтік хаттар, өтініш, анықтамалар, мінездеме, өңір тұрғындары жазған көркем мәтін, өлең;

- 1) Фольклорлық мәтіндер [189, 593 б.].

Орыс тілінің диалектілік корпусында сөздердің лексикалық мағыналары көрініс таппаған, грамматикалық белгіленімдер берілген. Орыс корпусында негізгі корпустың өзінде сөздердің лексикалық мағыналары берілмейді. Диалектілік корпусты құрастырушылар диалектілік ішкорпуста диалектизмдердің лексикалық мағынасын диалектологиялық сөздіктерге сілтемелер беру арқылы жасауға болатынын, алайда интернет желісінде мұндай сөздіктер базасы салынбағанын айтады.

Орыс тілінің диалектілік корпусын жасаушылар оның концепциясын бір-неше параметр бойынша жасаған. Олар:

- 1) Диалектілік материалды таңдау және диалектілік корпусың репрезентативтілік критерийін анықтау;
- 2) Корпустағы сөйлеу континуумын жіктеу принциптері;
- 3) Мәтіндік фрагменттерді беру параметрлері;
- 4) Корпустағы диалектілік мәтіндерді беру формалары;
- 5) Диалектілік мәтіндердің метабелгіленім параметрлері;
- 6) Корпусың мәтіндік базасын аннотациялау ережелері мен түрлері;
- 7) Диалектілік корпустағы экстралингвистикалық ақпараттардың берілуі;
- 8) Қолданушы сұранысына сай диалектологиялық зерттеулер мүмкіндігін жасаудың тиімді жолдары [190].

О.Крючкова көрсеткен параметрлер бойынша, диалектілік корпусың репрезентативтілік критерийі, ол – ұлт тілінің территориялық таралымын максималды түрде қамтылуымен анықталады. Сонымен қатар диалектілік мәтіндерді корпусқа енгізу нормаланған мәтіндерден гөрі анағұрлым қиын. Өйткені ғалымның айтуынша, диалектілік корпус диалектолог мамандардың қатысуымен, яғни белгілі біл говорды зерттеуші ғалымдардың қатысуымен жүзеге асырылады және мұнда аудио жазып алу, оны қағазға түсіру, қол жұмысы арқылы белгіленімдер енгізу сияқты қиындықтары мол жұмыс. Диалектілік корпуста репрезентативтілік принципі сақталу үшін жергілікті өңірлердегі барлық тілдік қарым-қатынас жанрларын, тақырыптарын қамту керек, яғни ұлт тіліндегі маңызды диалектілік типтер болу керек және өңірлік бөліністер бойынша теңгерімді материалдар корпус базасына салынуы қажет.

Мәтіндік фрагменттерді диалектілік корпусқа салуда да проблемалар кездеседі. Аудиожазба диалектілік сөйлесім фрагментін толық қамти ма әлде қысқа фрагменттерін қамти ма? Мысалы, Орыс тілінің ұлттық корпусында 3-4 сөйлемнен тұрады. Ғалымдар диалектілік корпус мәтінінің кіші бірлігі – абзац, үлкен бірлігі – тұтас аудиожазба мәтін деп санайды. Сонымен қатар диалектілік корпустағы аудиожазбалар стенограммасын орфографиямен не орфоэпиямен (транскрипция) беру мәселелі де үлкен проблема. Орфографиямен жазу ол мәтіндерді ортақ морфологиялық анализатор арқылы талдауға мүмкіндік берсе, екінші жағынан сөйлеу тілінің фонетикалық ерекшеліктерін тануға кедергі келтіреді дейді ғалымдар. Жазба стенограммасын аудио не видеомәтіннің символдық көрінісі деп атайды. Орыс ғалымдарының көрсетуінше, Орыс тілінің ұлттық корпусының Диалектілік ішкорпусында таспа жазбаларын символдық формаға келтіру әртүрлі диалектологиялық орталықтарда әртүрлі жүзеге асырылған. Ал Саратов диалектілік корпусында арнайы нұсқаулық бойынша қағазға түсіріледі. Онда ауызша жазба (дыбыстық) мен символдық жазба қатар орналасқан, ауызша мәтіннен жазбаша мәтінге, керісінше жазбаша мәтіннен ауызша мәтінге (аудиоға не видеоға) көшу мүмкіндігі жасалған.

Орыс тілінің және Саратов диалектілік корпусында лингвистикалық белгіленімдер негізінен әдеттегі негізгі корпустағыдай морфологиялық деңгейде жүргізіледі. Саратов корпусындағы тақырыптық және жанрлық белгіленім-

дер келешекте зерттеушілер үшін диалектілік сөйлеу тілін белгілі бір жанр, тақырыптық топтар бойынша зерттеуге мүмкіндік береді.

Орыс тілінің диалектілік корпусында метамәтіндік белгіленімдер:

- Жазба мерзімі;
- Жазба орны;
- Эксплотатор;
- Жалпы тақырыбы;
- Мәтін көлемі.

Крючкованың пікірінше, Орыс тілінің диалектілік корпусындағы бұл метамәтіндік белгіленімдер жеткіліксіз, аудиожазба жазылып алынған жағдаяттар (өткен жері (бақшада, орманда, үйде, мектепте т.б.), сөйлеу адресаттары, жазба фрагменттің жанрлық-тақырыптық құрылымы, мәтін ішінде баяндалатын оқиғаның мерзімі т.б. ескерілуі тиіс. Сонымен қатар диалектілік корпуста экстралингвистикалық материалдар ретінде көптеген қосымша ақпараттар салу қажет. Мәселен, фотосуреттер, географиялық карталар, тарихи, этнографиялық материалдар т.б.

Диалектілік корпус мәтіндік базасы жасалғаннан кейін 1) оны пайдаланушыға тауып алу, іздеу мүмкіндігін жасау және 2) ғылыми-зерттеу және т.б. дереккөздік мүмкіндігін қарастыру маңызды. Сонда ғана Ұлттық корпустар, сол сияқты Диалектілік корпус ғылыми-зерттеу жүргізудің теңдесі жоқ дереккөзіне айналады.

Қазақ тілінің ұлттық корпусына да осындай диалектілік мәтіндер ішкорпусын жасау – өзекті мәселе.

ҚТҰК-да қазіргі уақытта әдеби тілдің 5 стилін қамтитын Негізгі корпусы жасалған. Оның мәтіндік базасының көлемі қазіргі уақытта – 40 миллионға жуық сөзқолданыстан тұрады. Онда әр стильге тән метамәтіндік белгіленімдер салынған.

ҚТҰК-ның Диалект мәтіндер ішкорпусы әзірге «Қазақ тілінің аймақтық сөздігі» материалдары негізінде жасалды. Ол үшін:

– біріншіден, «Аймақтық сөздіктің» реестрлік сөзтізбесі Диалектілік ішкорпустың негіз сөздер базасы ретінде корпусқа енгізілді;

– екіншіден, әрбір диалектизмге берілген сөз мағынасын дәйектейтін мысалдар Диалектілік ішкорпустың мәтіндік базасы ретінде корпусқа салынды, яғни мәтін көлемін «Аймақтық сөздікте» берілген мысалдар құрайды. Сонымен қатар онбестомдық «Қазақ әдеби тілінің түсіндірме сөздігінің» иллюстрациялық материалдары қосылды. Жалпы көлемі – 3 294 049 сөзқолданыс.

– үшіншіден, «Аймақтық сөздіктегі» әрбір сөздің өңірге қатысы тіркеліп, бұл Іздеуіште (Поисковик) «Өңір бойынша іздеу» жасауға мүмкіндік берді;

– төртіншіден, реестрдегі әрбір сөзге сөз табы белгісі қойылды.

Диалектілік корпустан ізделген сөздің лингвистикалық терезесінде алдымен олар түрленген тұлғада тұрса, негізге (негізгі және туынды түбір) келтіріліп, яғни леммасы бөлініп, ол негізге сөз табы белгісі қойылады.

Жергілікті тілді зерттеушілер үшін корпуста жазылып алынған дискурстық мәтіндердің жанрлық түрін, тілдесімге түсушілердің жас, жыныс, әлеуметтік (білімі, қызметі т.б.) ерекшеліктерін білу маңызды. Осы орайда диалектілік мәтіндер корпусында жергілікті тілдесімдерді жинауда ол тілдесім

Диалектілік мәтіндер ішкорпусындағы сөз терезесінде метабелгіленімдер мен лингвистикалық белгіленімдер осылайша аралас берілді.

Ал диалект сөздер алынған мәтіндер туралы ақпараттар, жоғарыда атап көрсеткеніміздей, 2-ақ дереккөзден алынды («Аймақтық сөздік» және Қазақ әдеби тілінің сөздігі») және метамәтіндік белгіленім терезешесі осы 2 дереккөз бойынша рәсімделді.

Диалект мәтіндер ішкорпусының Іздеу жүйесі 3 параметрмен жұмыс істейді. 1 – Сөзформадан іздеу; 2 – Сөзден іздеу; 3 – Өңір бойынша іздеу:

Бұл жоғарыда берілген жұмыс үстелі – Қазақ ұлттық корпусы Диалектілік ішкорпусының бастапқы нұсқасы болып табылады. Мақала басындағы Әдебиеттік шолуда басқа тілдердегі Диалектілік корпусстарға талдау жасалған болатын. Диалектілік корпус жасаудың әлемдік тәжірибесін осылайша әлі де зерттей түсіп, Қазақ тілінің диалектілік ішкорпусының әзірлемесін жетілдіре түсуіміз керек. Бұл ретте алдағы уақытта бірнеше бағытта жұмыстар ұйымдастырылу керек. Олар:

– Диалект мәтіндер ішкорпусының мәтіндік базасын өңірлерге тән ауызша сөйлеу тілінен аудиотаспаға жазып түсіру қажет, ол үшін диалектологиялық экспедициялар ұйымдастыру керек;

– Жазылып алынуға тиіс жазбаларды жанры мен тақырыбы бойынша жобалау қажет, яғни филологиялық сараптама жасау керек, бұл корпусстың репрезентативтілігін қамтамасыз етеді;

– Жазылып алынатын аудиотаспалар жазу кезінде қандай параметрлер бойынша тіркеп алу керектігін жоспарлау қажет, яғни аудионы жазып алушы неге мән беру керек, нені тіркеп отыру қажеттігін білуі керек;

– Жоспарланған параметрлер тіркелген аудиожазбаны қағазға түсіру керек, бұл үшін алдымен ауызша сөзді орфографиямен не орфоэпиямен (фонетикамен) жазу туралы шешім қабылдануы керек және стенограммаға түсіру туралы нұсқаулық әзірленуі керек, сонымен қатар қазіргі кезде ауызша сөзді жазбаша нұсқаға автоматты түрде келтіретін программалар бар (сөйлеу тілін анализдеу және синтездеу), осы программаларды ауызша сөзді жазбашаға, жазбаша сөзді ауызшаға айналдыратын мүмкіндігін бақылау және оны Ауызша корпусы, Диалектілік корпусы жасауда пайдалану қажет;

– Жазылып алынған ауызша мәтіндердің метаақпараты метабелгіленім терезесіне (мәтін паспортына) тағы да бірнеше жаңа параметр қосуды қажет етеді, мысалы, сөйлесушілер туралы ақпараттар (кімдер, жасы, жынысы, білімі т.б.), қашан, қайда жазылғаны және сөйлесім тақырыбы, жанры (БАҚ, тұрмыстық сөйлеу, іскери сөйлесім, т.б.).

Қазақ тілінің диалектілік мәтіндер ішкорпусын диалектолог-тілші-ғалымдар, басқа да лингвист-ғалымдар, филологтар, тарихшылар, мәдениеттанушылар, этнографтар, басқа да лингвист емес пайдаланушылар әртүрлі мақсатта қолдана алады. Диалектілер мәдениеттің бір бөлігі екені даусыз. Жергілікті халық диалектілерінің осындай маңызын ескере отырып, диалектілік корпусын құру өзекті мәселелердің бірі болып табылады. Жыл өткен сайын ғылыми зерттеу жұмыстарын жүргізу үшін толық қамтылған ақпарат көздерін цифрландыру қолға алынып келеді.

Диалектілік корпус мәтіндері мен онда қамтылған түрлі экстралингвистикалық және лингвистикалық белгіленімдерді қолданушылар ғылыми және педагогикалық жұмыстар материалы, этнографиялық, этномәдени дәстүрлер, қазақ менталитетінің, қазақ жергілікті тілінің ерекшеліктерін көрсететін анықтамалық құрал, дереккөз ретінде пайдалануына болады. Диалектілік корпус стандартқа сай жасалғанда ғана диалектологиялық зерттеулер жүргізудің негізгі дереккөзі болу қызметін атқара алады.

8 МӘДЕНИ РЕПРЕЗЕНТАТИВТІ МӘТІНДЕР ІШКОРПУСЫ

8.1 Этнолексика – ұлт болмысының тілдік дерегі

Қазақ тілінде қалыптасқан этнолексика – тіл арқылы ұлттың болмыс-бітімін, дүниеге көзқарасын, рухани-мәдени құндылықтарын, әлеуметтік мәнін жинақтаған танымдық-ақпараттық көз. Ол халықтың этномәдени тарихы мен тілдің шығармашылық даму сабақтастығын, тілдің кумулятивтік және коммуникативтік қызметін айғақтайтын сенімді дәйек. Осы тұрғыдан оның айшықты көрінісі тарихи даму барысындағы тілдік жүйенің шеткері өрісін құрайтын фразеологиялық құрылымдарда сақталған. Ғасырлар бойы халықтың көкірегінде жатталып, жадында сақталған, кестелі сөз шеңберінде бейнеленіп, ұрпақтан-ұрпаққа жеткен көне тамырлы сөз ретіндегі фразеологизмдердің мазмұнында этномәдени код астарланған. Ассоциативті-бейнелі ойлаудың нәтижесі ретінде фразеологизмдер семантикасының этномәдени және экстралингвистикалық деректері ашылады. Белгілі бір этномәдени бұйымның атауы ретіндегі жеке сөздің қолданыстық мағынасының кеңеюі (мысалы, *талқы*, *тұсаукесер* т.б.) немесе тарылуы (мысалы, *тон* т.б.) тіл иесінің санасымен, әлеуметтік мәнімен байланысты. Тіл құдіреті арқылы сақталған осындай мазмұндық дерек – сол тілді тұтынушы этностың бүкіл рухани-мәдени байлығының коммуникативтік қызметінің негізінде бүгінгі заманға сай әлеуметтік өрісі кеңейіп, танылып, жалғасуда.

Сайып келгенде, көне тамырлы этнолексика – халықтың этномәдени тарихы және тілдік даму көріністері мен тілдік шығармашылығы туралы баға жетпес «ақпараттық» және танымдық көз. Оның айшықты көрінісін фразеологиялық тіркестердегі этнотаңбалардан көреміз. Мысалы, *байғазы беру*, *қалжа жеу*, *қанжысаға байлау*, *бесік құда*, *ежегабыл құда*, *киім кию* т.б. Осындай этнотаңбалық қызметіне орай фразеологизмдерді көркем сөз бен шешендік өнерінің үлгісі ретінде тануға болады. Олар тілтұтынушының санасындағы тарихи-мәдени, рухани-әлеуметтік т.б. дүниетанымдық әлемін танытатын этнотаңбалық тілдік бірліктер ретінде дамып жалғаса береді. Сөйтіп, профессор Қ.Жұбанов көрсеткендей, олар тілдің коммуникативтік қызметімен сабақтас кумулятивтік қызметі негізінде тіл құрылымының негізгі арқауын жалғап, тарихи дамуы барысындағы тілдік жүйенің шеткері өрісін құрайтын құрылымдарда сақталады. Сондықтан ғасырлар бойы халықтың көкірегінде жатталып, жадында сақталған, көркемсөз кестесінің шеңберінде айшықталып, ұрпақтан-ұрпаққа тілдің коммуникативтік те, кумулятивтік те қызметі негізінде жеткен көне тамырлы сөздердің негізгі көзі, этномәдени коды – фразеологиялық тіркестер. Олар – ұлттық болмыс-бітімін, дүниеге көзқарасын, рухани-мәдени құндылықтарын, әлеуметтік мәнін, ой тәжірибесін, даналығын, дүниетанымдық тұжырымдарын, салт-дәстүрін ұлттық таным қорында жинақтаған тілдік таңбалар, танымдық мазмұны астарланған сөздік ақпарат. Жеке сөздердің тура мағынасындағы кейбір көмескі тұстарының толық ашылуы фразеологизмдердің қосымша мағыналық қызметі арқылы жүзеге асады.

Фразеологиялық құбылыстың мәнін тілдің номинативтік жүйені толықтыруға қатысымен байланысты түсіндіруде ғалым С.Исабеков өз зерттеуінде «принцип дополнительности» ұғымының ерекше маңызын көрсетеді [191]. Соның нәтижесінде кеңітіліп бейнеленген мазмұн өзінің суреттілігімен, астарлы мағынасымен экспрессивтік эмоционалды әсерді күшейтеді. Себебі осындай ұлттық бояуы қою, көркемдік сипаты жоғары айшықты тілдік құралдар ретінде фразеологизмдерде мәдени тілдік ұжымның байырғы дүниетанымы, салт-дәстүрі жеке сөздерге қарағанда айқын көрінеді.

Фразеологизмдер семантикасының таңбалық мәртебесін ғалым Р.А.Авакова этномәдени және экстралингвистикалық дереккөздері арқылы ашып, олардың таңбалық-символдық белгілерін ассоциативті-бейнелі ойлаудың жемісі ретінде анықтайды [192].

Зерттеуші ғалым Н.Уәли *тайға таңба басқандай, қой үстіне бозторғай жұмыртқалаған заман т.с.с.* идиомалар көркемдік ойлау деңгейін ғана емес, ұлттың күнделікті тұрмыс тәжірибесінен қалыптасқан танымдық тұжырымын да белгілейтінін атап көрсетеді [193].

Сол сияқты қазіргі қазақ тілінде *талқы* сөзі *талқыға салу, талқыға түсу* сияқты тіркестердің немесе *талқылау, талқылану* деген етістіктердің құрамында қолданылады. Ал тарихи тұрғыдан **талқы** – «*теріні керіп созып, иін қандыра жұмсартатын ағаш аспап*» (Қазақ сөздігі, 1212-б.). Демек, тарихи тұрғыдан құралдың атауы ретінде қолданылған жеке сөздің қолданыстық мағынасының кеңеюі тіл иесінің санасымен байланысты (*өмірдің талқысы, тағдырдың талқысы, талқыға түсті, талқыға салды, талқы көрді т.б.*) туған. Бұл сөздің бастапқы мағынасы, мысалы, «Қозы Көрпеш – Баян сұлу» жырында көрініс тапқан:

Сарттар сауыр басады талқыменен,

Ер жабдығы жарасар нарқыменен.

Жетім ұлға қызымды бермеймін деп,

Қарабай көшіп кетті халқыменен.

Ескі түркі жазбаларында **талқы** сөзінің етістік тұлғасының кездесуі бұл сөздің синергиялық көзі *tal* түбірінен тарауды деп жорамалдауға дәйек береді: *talqi* – «*кінәлау, қабыл алмау*»; *talqiš* – «*киіз басу*» [194, 529].

Соған сәйкес халқымыздың рухани мәдениетінің тілдік бейнесі эпостық жырларда, көне жазба ескерткіштерде, ақын-жыраулар мұраларында, мәдени-құжаттық сипаттағы атауларда, мифтік, прецеденттік, паремиологиялық, фразеологиялық, этнографиялық т.б. этномәдени контексте жаңғыртылады.

Синергиялық желі тізбегінде жаңғырылып жалғасқан олардың мазмұны күнделікті тұрмыс пен өмір тәжірибесін таңбалайтын рухани дүниесінде этнопедагогикалық, экономикалық, экологиялық, құжаттық т.б. дағдыларын, өзіндік дүниетанымы мен құқықтық, қоғамдық этика нормаларына, өмір тәжірибелеріне қатысты құндылықтарын қалыптастырып, тіл арқылы сақтайды: *ала жіпті аттамау, ат құйрығын кесісу, бетегеден биік, жусаннан аласа, құйрық-бауыр жесу т.б.*

Мысалы, *құйрық-бауыр жеу* тіркесі елдің бірлігін, туыстығын, ымыралауын сипаттап, екі жақты тұтастыруды мақсат ететін рухани мәдениеттің тілдік көрінісі.

Қазақ тілінің он бестомдық сөздігінде *құйрық-бауыр жеу* тіркесіне мынадай түсіндірме берілген: «Қазақ халқының келін түсіріп, қыз ұзатып құдаласуы барысында орындалатын тәрбиелік мәні зор дәстүрлердің бірі – **құйрық-бауыр жеу**. Құйрық-бауыр жеу қазақ дәстүрінде құда болудың заңды белгісі – құжаты болып бекітілген» (Қазақ сөздігі, 836). Бұл тіркестің этномәдени мазмұнының уәжін белгілейтін ата-бабаларымыздың ата кәсібі және сонымен байланысты күнделікті тұрмыс-тәжірибесінде дәлелденіп, символдық сипатқа ие болған деректердің коннотаттық сипаты. Рухани мәдениеттің мұндай архетиптік бейнесі ретіндегі сөз үлгілері тілдік қазынамыздың этномәдени қорында жетерлік. Мысалы, *Ер Төстік* антропонимі немесе *шаңырақ, қазан, қамшы* т.б. онимдік бірліктер. Салттық мәдениеттің елдік, этностық белгілерінің көбі жеке сөз, ия сөз тіркестері түріндегі тілдік бейнесінде сақталған: *тұсаукесер, естірту, үкі тағу, иткөйлек* т.б.

Қазақ танымында қасиетті құс деп танылған үкінің «тіл-көзден, пәлежаладан, жамандықтан сақтайды» т.с.с. киелі қасиеттері **«үкі тағу»** тіркесінің мифтік танымға сәйкес мағынасы этноәлеуметтік қызметімен өзектелгенін этнограф-ғалымдар да атап көрсетеді [195, 576-577]. Осы мәнімен байланысты **үкі тағу** ұғымы Қазақ әдеби тілінің он бестомдық түсіндірме сөздігінде былайша анықталған: «Күйеу жігіттің әке-шешесі, туыстары құдаласқан күні қалыңдықтың бас киіміне үкінің жүнін тағып, қызды айттырып қойғанының белгісін жасауы. Үкі тағылған қыз сол күннен бастап айттырылған болып, оған басқалар құда түспейді (Қазақ сөздігі, 1347).

Жоғарыда көрсетілген *үкі тағу, иткөйлек* сияқты тілде сақталған рухани мәдениеттің этнотанбасының уәжін олардың мифтік дүниетанымға сай тотемдік мәнімен байланыстырсақ, ұғымның өзегі ретіндегі тілдік деректерді халқымыздың ұзақ этномәдени даму жолының материалдық және рухани құндылықтарының түп-тамырын неғұрлым тереңірек танудың бір құралы, маңызды таным тетігі деп тану тіл мен таным, тіл мен әрекет сабақтастығының ашық жүйе екендігін дәлелдейді. Мысалы, *иткөйлек* этнолексемасының мәдени семантикалық уәжі мифтік таным деңгейімен жалғасады. *Иткөйлекті* жаңа туған бала қырқынан шыққанға дейін кигізеді. Мифтік танымға сай **ит** – о дүние әлемінің өкілі. Демек, бала қырық күннен кейін қарақұлақтанып, ширап осы дүниенің әлеміне енеді деген қазақы дүниетаным қалыптасқан.

Қазіргі қазақ қоғамында халық рухын, оның өзегі мәдениетті тіл арқылы жаңғыртудың бір үлгісі қазіргі күнде жаңа терминдік мағынаға ие болған *тұсаукесер* сөзі. Бұл сөздің ұлттық термин ретінде қызметі ұлт санасының тереңіндегі оның архетиптік мағынасының қазіргі замандағы тіл иесінің санасында ассоциациялық жаңғыруының нәтижесінде жүзеге асып отыр. Шын мәнінде, қазақтың төл дәстүрінде **тұсаукесер** ұғымы «Сәби қаз тұрған соң тез жүріп кетсін деген тілекпен жасалатын ғұрып, ырым» деп түсіндіріледі. Ал қазіргі тілдік қолданысқа енген «алғаш рет таныстыру рәсімін» білдіретін терминдік мағынасы тілдің өміршендік, мұрагерлік қызметін дәлелдейді.

Әрине, мәдениетті, рухани көріністерді бейнелейтін тіл, миф, өнер, дін және т.б. семиотикалық жүйе құрады. Бірақ қазақтың көшпенді мәдениетінде рухани мәдениеттің таңбалық негізі, рәміздік жүйедегі ұлттың психологиясы, сакрализациясы негізінен тіл арқылы сипатталады. Сондықтан рухани мә-

дениеттің коммуникативтік, мұрагерлік табиғатына сәйкес сақталуы, кейінгі ұрпаққа жетуі, жаңғыруы, танымы игерілуі, жетілуі оның ашық жүйе ретінде тіл арқылы іске асатыны, тіл иесінің коммуникативтік болмысымен байланыста ашылуы – заңдылық. Соның негізінде бір этностық (ұлттық) ұжымда мәдени коммуникацияның іске асырылу құралдарының дені, таңбалар жүйесі, рәміздер, әдет-ғұрыптар, материалдық өндіріс туындыларының өзі тіл арқылы мәдениет тарихи-элеуметтік үдеріс ретінде сипатталады. Сондықтан да қазіргі замандағы саналы ұрпаққа қазақ этносының таным-парасатын, дүниетанымы мен мәдениетін тіл арқылы таныту, атап айтқанда, мәдени семантикасын ұлттық корпуста ашып беру – аса маңызды.

Мысалы «**Жеті қарақшы**» космоимдік атауына жеті санының киелі мәні мен прецеденттік мәтін уәж болып тұр десек, тіліміздегі *жеті нан* деген діни мазмұнды қолданыс пен табиғаттан тыс ерекше күшті бейнелейтін *жетібасты жалмауыз* сияқты мифтік атаулардың қалыптасуы тіл иесі танымындағы жеті санының символдануымен түбі бір тамырда сабақтасады. Қазақ танымына сай төл мәдениетте бұл санның қалыптануы мен заңдасуының бір дәлелі – қазақтың «жеті атадан бері қыз алыспау» заңы. Күнделікті тұрмыс пен өмірдегі көп нәрсені осы өлшеммен белгілеу де қазақ танымына сай қалыптасқан: *жеті атасынан шынжыр балақ, шұбар төс; жеті атасынан бері қонған (ақындық, ұсталық, сынықшылық т.б.)*. Бұл арада *жеті* сөзі өзінің тура мағынасынан асып, жалғасып «о бастан», «ежелден», «арғы аталарынан» т.с.с. ұғымды бейнелейді. Сол сияқты «жеті жұрттың тілін біледі» дегенде де тура «жеті» мағынасын емес, «кең» мағынаны меңзейді.

Демек, тіл – мәдениет дерегі ретінде мәдениеттің құрастырушы бөлігі және мәдениетті танудағы негізгі құрал, сондықтан ол арқылы біз мәдениетті меңгереміз. Себебі тіл өз мазмұнында мәдениеттің жалпы сипатын танытады, негізгі этномәдени ақпаратты жинайды, сақтайды және жеткізеді. Осыған сай тіл – әрбір этникалық қоғамдастықта, бір жағынан, этностың мәдениетін сипаттап, сақтаушы ретінде, екінші жағынан, тіл шеңберінде мәдениеттің ұрпақтан-ұрпаққа жалғасуын, келер ұрпақты бір тұтастырып, бүтіндікте байланыстыратын қуатты коммуникативтік құрал.

Адам санасында қалыптасатын ақиқат дүниенің бейнесі арқылы сақталатын тілдік қорда әрбір элеуметтік жүйенің, ұлттық ұжымның, әрбір қоғамның өркениеттік даму барысына сай қалыптасатын дүние бейнесі тілде өзіндік ерекшеліктерімен ажыратылып отырады. Олардың негізінде жеке адамның танып түсіну деңгейі мен дәрежесі бара-бара ұлттық сипатта тұтасуы тілдік санада сезімдік қабылдаулар нәтижесінде концептуалдық жүйе арқылы жүзеге асады. Осыған сай дамып, дүниенің қасиеттері одан әрі таныла бастаған сайын дүние бейнесінің мазмұны да өзгеріп отырады. Оның мысалын жоғарыда көрсетілген *тұсаукесер* сөзінің бастапқы этнографиялық ұғымының қазіргі таңдағы терминдік сипатқа ие болуынан көреміз. Ол тіл арқылы ұлт болмысын тануға ұмтылған қазақтың бүкіл заттық және рухани, мәдени атаулары мен ұғымдарының қазіргі ұлт санасында этномаркерленіп, жадыда терең сақталу құбылысын көрсетеді. Мысалы, *ежеғабыл құда, бел құда, бесік құда, жаушы, қалың мал, бес жетім, сауын айту т.б.*

Шын мәнінде, таным процесі мен логикалық ойлау жалпы адамзатқа ортақ қасиет, ал бүкіл адам баласы үшін ақиқат шындық болмыс біреу болғандықтан, ғаламның біртұтас бейнесі де әр басқа тілде сөйлейтін халықтар үшін біреу-ақ. Сөйтсе де тіл иесінің ой-өрісіне, концептуалды әлеміне байланысты әр тіл дүние-ғаламның фрагменттерін өзінше өреді, бөлшектеп атайды, бейнелеп суреттейді, ал тіл өкілі ойының желісі өз тілінде бекіген ғалам бейнесіне сай құралады. Осыған байланысты әр басқа халықтың (ұлттың, этностың) тілдеріндегі ғалам бейнесінің фрагменттерінде жалпы адамзатқа ортақ дүниеліктер де, тек тіл иесіне ғана тән ұлттық нақыштар та кездесіп жатады.

Демек, әрбір этностың мәдениеті – оның тілі арқылы сақталған «ішкі» мазмұнында. Ал, «мәдениетті тану кілті» – тіл арқылы сақталған мәдени ұғымдардың мәдени фонын анықтау десек, соның нәтижесінде анықталатын тілдің қызметі – әр халықтың ұлы мұрасы, ұлттың тарихи жетістігін көмкерген мәдениетті таңбалау.

Жалпы алғанда, мәдениеттің немесе оның жеке түрлері мен формаларының таңбалық сипаты туралы ой ертеде қалыптасқан. Ал тілге де қатысты **таңба** ұғымының тіл біліміне енуі тілден тыс салт-дәстүр, әдет-ғұрып т.б. мәселелерді *тіл* мен *мәдениет*, *ұлт* пен *тіл* қағидасына сәйкес танымдық сипатта қарауға мүмкіндік туғызды. Сондықтан тіл арқылы мәдениет мазмұнын ашуға қатысты қызмет атқаратын таңбалар жиынтығы мен таңбалық құралдарын **этнотаңбалар** (Ю.М.Лотман т.б.) немесе **этномаркерлі лексика** деп атау оларды лингвистиканың антропоэзеттік парадигмасына енгізеді.

Жалпы алғанда, мағыналық таңба белгілі бір ұғымның ерекшеленген қасиеті, қызметтік белгісі, бейнесі арқылы айқындалып, мағыналық мәні таңба ретінде жадыда қалыптасып, сөз қолданысында заттық таңба ретінде ұғымды танытатын әлеует деңгейін сақтап қалады. Осыған орай ең негізгі құрал – сөздік ақпарат. Оның негізін құрайтын этнографизмдер, қазақ ескіліктері, мифологемалар, басқа да этномәдениеттің архетиптік көріністері мәдениетті айқындап қана қоймай, оны ұлт болмысымен тұтастыратын символдау қызметін атқарады да, соның негізінде этнотаңбалы мәдени кеңістік құралады. Оның тілдік көріністері ретіндегі лингвокультурема, этнографизмдер мен одан туындайтын тұрақты тіркестердің түп тамыры ұлт санасының терең құрылымдарында тұрақты түрде сақталып, сөз мазмұнындағы этномаркерлік деңгейге жеткен таңбалардан бастау алады. Сөз мазмұнында өздерінің іштей ұйымдастырылу, реттелу т.б. синергиялық үдерістерінің нәтижесінде олар ақпаратты жеткізуші, қабылдаушы, сипаттаушы, белгілеуші, бағалаушы, жалпылаушы сияқты коммуникативті-прагматикалық қызметтерді атқарады.

Осындай қызмет атқаратын таңбалар қатарына символдар, бейнелі сөз-образдар, эпитеттер, метафоралар, метонимия, т.б. жатады. Дегенмен, ұлттық таным-түсінікте кейбір ұғымдарды астарлап, тұспалдап жеткізуде (*өлімді естірту, қаралы хабарды жеткізу т.б.*) қолданылатын шартты, бейвербалды таңбалар да өз мәнін жоймай, бүгінгі күнге дейін ұлт жадында сақтаулы. Мысалы, *қара ту көтеру, атын бос жіберу, ауылға шауып келу, қамшыны ортаға тастау т.б.* Осыған қатысты В.А. Маслованың еңбегінде символдардың әмбебап түрлері мен ұлттық түрлері және басқа таңбаларға қарағанда

символдың бойында терең ұлттық мән болатыны туралы пікірі маңызды [196]. Соның негізінде ұлттық сипаттағы таным-түсініктердің тілдік айшықталуы символдардың қолданысынан айрықша көрінеді. Олардың танымдық уәждері де сан алуан. Мәселен, *Айдың аманында, Айы оңынан туды, Ай тұяқ шалды* деген тіркестердегі **ай** символының қызметі мен *Жұлдызы қарсы, Жұлдызы теріс келді, Жұлдызы қарсы, Жұлдызы жоғары* дегендегі **жұлдыз** символының қызметі наным-сенім, сыйыну тотемдік, шамандық дәуірден бастап ұлттық санада сақталып, генетикалық тұрғыдан жалғасып келе жатқан мифологемдік этномаркерлер қатарына жатады.

Ал ұлттық заттық-мәдени лексемалардың этномәдени сипатқа ие болып, күнделікті тұрмысты, ата кәсіпті, мінез-құлықты, іс-әрекетті, құбылысты бейнелеуде тұрақты қолданылу нәтижесінде стереотиптік этномаркерлер пайда болады. Мысалы, *айрандай аптап, күбідей күптеу (толықтай өзіне қарату, бағындыру); ақ киізге көтеру; ақ түйенің қарны жарылу, Асан қайғыға түсу; асығы ашысынан түсу; асығы түгел; аузынан ақ май ағу; ауызды қу шөппен сүрту; аузыңа май, астыңа тай; ауыл итінің құйрығы қайқы; әңгір таяқ ойнату; бармағын тістеу; бас жарылса, бөрік ішінде, қол сынса, жең ішінде; безбұйрек; бетегеден биік, жусаннан аласа; бетке күйе жағу, бетке таңба болу; бет моншағы үзілу, бетін айдай қылу; бие сауымдай уақыт; бір күн бие, бір күн түйе; бір тарының қауызына сыйғызу; дәм-тұз атсын; дәм-тұзы жарасу; екі қошқардың басы бір қазанға сыймайды; есік көрген; етінен өтіп сүйегіне жету т.б.*

Сонымен, ұлттық дүниетанымның нәр алатын ортасы, әрі көрінер жері – мәдени коммуникация. Мәдениеттің мазмұнында халықтың мыңдаған жылдар бойындағы шығармашылығы, қауым мен жеке адамның рухани ізденісі, халықтың даналығы мен адамгершілік нышандары жинақталады. Соның негізінде әр мәдениеттің бойында дүниеге, қоршаған ортаға деген сан қилы көзқарас, түсінік қалыптасады. Себебі әр мәдениет заманына, діліне, моральдік нормаларына қарай адам феноменіне, өмір – өлім ұғымдарының философиялық дүниетанымына, адамның іс-әрекетіне, сұлулыққа тағы да басқа әр халықтың басында бар негізгі құндылықтарға байланысты өзінің көзқарас жүйесін қалыптастырады. Міне, осы көзқарастардың көрініс тауып, таралатын, сақталатын құралы – тіл.

Әрбір этностың ежелден қалыптасқан төл мәдениеті – оның тілі арқылы сақталған, таңбаланған ішкі мазмұны. Сол мазмұнды ашу арқылы ұлт мәдениетінің мәні айқындалады десек, «мәдениетті тану кілті» сөздің синергиялық жүйесінде құрылымдалған мәдени ұғымдардың фонында астарланған, тіл құдіреті арқылы берілген мазмұндық дерек, сол тілді тұтынушы этностың бүкіл рухани, мәдени байлығының асыл қазынасы.

Тілдің осындай мұрагерлік (кумулятивтік) қызметінің бүгінгі заманға жалғасқан әлеуметтік өрісін кеңейтіп, ұлттық мазмұнының танылуын ұрпақтан ұрпаққа жалғастыратын жаһандану заманына сәйкес кешенді танымдық кеңістік, заманауи тиімді ұтқыр құрал болатын А.Байтұрсынұлы атындағы Тіл білімі институтында жасақталып жатқан Қазақ тілінің ұлттық копрусы. Қазіргі таңда бұл жетістіктің тиімділігі мен құндылығы қазіргі қоғамның рухани әлеуметтік даму мен жаңғыру үрдісінің талап нүктесімен үндесуінде.

Уақыт пен кеңістік сабақтастығында қарастырылатын кешенді лингвомәдени зерттеу барысында бір кезде күнделікті өмірде өте қажетті болған көптеген заттардың атаулары ғана қазіргі қазақ тілінің қазынасында мәдениеттің, дәстүрдің және ұлттық мұраның жұрнағы ретінде сақталып қалғаны анықталады. Осыған орай Ұлттық корпуста олардың бойындағы ұлттың рухани қазынасын бойына сіңірген құжаттық, этномәдени таңба ретіндегі қызметін қазіргі тілдің негізгі коммуникативтік қызметі негізінде мәдени семантикасын жаңғырту елдің мәңгілік құндылықтарының нақты мазмұнын ашуға негіз болады.

8.2 Ұлттық корпуста ұлттық мәдениеттің этномәдени кодтарын ашу мәселелері

Тіл білімі – ғылымның дербес бір саласы. Ғылым, қандай да болмасын, ғылымның дербес саласы болу үшін оның қарастыратын объектісі, предметі, зерттеудің әдіс-тәсілдері болуға тиіс. Тіл ғылымы ғылымның дербес саласы ретінде әуелде бірден қалыптаса қоймағаны белгілі. Тіл ғылымының негізгі зерттеу объектісі – тіл екені еш дау тудырмайды. Бірақ, мысалы, металл физиканың да, химияның да объектісі болып табылатыны сияқты тіл философияның да, психологияның да, физиологияның да объектісі болуы ықтимал. Бірақ олардың тілді зерттеу қыры әр басқа, яғни предметі мүлде басқа. Физика тілдің дыбыстық жағын оның ауа толқыны арқылы сипаттайды. Тілдік дыбыстың күшін, қарқынын өлшеп белгілейді, тілдік дыбысты физикалық құбылыс ретінде ғана таниды. Ал оның мағына ажырататын жағын, айырым белгілерін тексермейді.

Адамзат қоғамы мәдениеттің небір кезеңдерін басынан өткерді: атап айтқанда, адамдар еңбектің қол құралдарын жасады. Сол құралдармен егін екті, мал бақты; заттық мәдениеттің түр-түрін қол құрал күшімен жасады. Сонымен қатар өмір тәжірбиесінен өндіргіш күшке қатысты білім жинақтады. Бұндай білімдер жүйесі рухани мәдениеттің дамуына да негіз болды. Мәдениеттің өркениет кезеңінде адамзат қоғамы еңбек құралдарының түр-түрлерін жетілдіре бастады. Осымен байланысты адамдардың да тұрмысы мен санасы өзгерді, еңбек өнімділігіне қол жеткізді. Келе-келе машиналар заманы туды. Адамдар дене күшімен, яғни қолмен (қайла, күрекпен) атқарылатын жұмысты машинаға «тапсырды». Ал адамның өзі машине басқарушылық типке айналды. Енді осыған сәйкес қоғамның әлеуметтік құрылымы өзгерді: жетілген техника заманында он, жиырма, жүз, мың адамның жұмысын машина басқарушы бір адам ғана атқарды. Күрек-кетпеннен босаған адам қоғамдағы әр алуан әлеуметтік саланың мамандары ретінде дәрігер, мұғалім, механик, инженер және т.б. болып, игілікті істер атқарды.

Адамзат мәдениетінде адами-қоғамы, өзгерістер күшейіп, ғылыми-техникалық прогресс барған сайын үдей түсті. Бұл кезеңнің ерекшелігі сол – мимен істелетін, миға күш салатын жұмыстарды электронды есептеуіш машиналар, суперкомпьютерлер атқарып, әлемді ғаламтор шырмай бастады.

Уақытты көп алатын ми жұмысына компьютерлер «киліккен» соң, адамдардың еңбек уақытынан тыс «компьютерлік бос уақыты» көбейді. Ғаламтор

торабында бірінен-бірі өткен небір бос уақытты толтыратын «бағдарламалар» көбейді. Ұялы телефондар интернетке қосылды. Осылайша «миы компьютерленген адамдардың» типтері пайда болды. Ондайлар айналасында не болып жатқанына бейжай қарайтын, өз қызықтары өздерінде болып, рухани әлемнен оқшаулана түсті. «Компьютер адамның құлы болудан қалып, адам компьютердің құлына айналды ма?» деген күдікті ой – заманның, уақыттың сұрауы.

Адамзат қоғамның заттық, материалдық игілікті өндіруі бұрын-соңды болмаған жедел дамудың даңғыл жолына түсті. Моральдық жақтан ескіріп, физикалық жақтан тозып үлгермей-ақ түрлендірілген саналуан заттық дүниелер тұтынушыларды шырмаған торға айналды. Бұл құбылыс экономикалық жағынан бір қалыпты тұрақты даму емес, тұрағы жоқ даму сияқты көрінеді. Сөйтіп, материалдық игілікті өндіру қарқыны үдей түскен сайын қоғамда әсіре тұтынушы адамдардың типтерін көбейте түскендей көрінеді.

Қазіргі қоғамның мәдени кеңістігінде эстетикалық санасы тобырлық мәдениетпен (массовая культура) суарылған әлеуметтік жіктің жылдан-жылға көбейе түскені байқалмай қалмайды. Сөйтіп, адамдардың тобырлық мәдени типі қалыптасып жатқаны да шындық.

Қысқасы, ұлтсызданған сана, әсіретұтынушылық сана, компьютерленген сана, тобырлық мәдени сана, түптеп келгенде, ұлттық сана-сезімнен алшақтап бара жатқан жаһанияттың жергілікті өнімдері деуге болады. Жаһанияттың бұндай өнімдері қай ұлт, қай ел болмасын тіл тәрбиесі осал, тіл білуі мардымсыз, тілге жеңіл-желпі қарайтын ортада өсіп-өнгіштеу келеді. Өзінің ұлттық сипатын, елдігін сақтау, ұлттық құндылықтар дүниесін түгендеу, оның мән-мазмұнын, қадір-қасиетін бағалау қай халықтың, қай елдің, қай ұлттың болмасын толғақты мәселесіне айналып отыр деуге болады.

Материалдық даму мен рухани дамудың арасы алшақтай түсуі бүгінгі таңдағы ғылыми зерттеудің күн тәртібіне қойылып отыр. Қоғамның рухани дамуы мен материалдық дамуы арасындағы қайшылық гуманитарлық-қоғамдық саладағы ғалымдардың осы мәселеге ерекше мән беруіне әсер етті.

Еліміздің білім беру ордаларында маңдайалды кәсіби маман даярлауға ерекше ден қойып отырған ғалым педагогтарды шәкірттерінің рухани мәдениетінің деңгейі алаңдата бастады. Осымен байланысты тіл өміршендігі (вита́льность языка), тіл бірегейлігі (языковая идентичность), ұлт бірегейлігі (национальная идентичность), этнос бірегейлігі (этническая идентичность) мәселелері гуманитарлық ғылымда өзекті тақырыпқа айналды.

Ұлттық корпуста мәдени кодты ашуда тіл құзыреттілігімен қатар мәдени-танымдық құзыреттіліктің болуы шарт. Лингвомәденитанымдық құзірет дегеніміз – тіл иелменінің мәдени-тілдік таңбадағы кодқа салынған ақпараттарды аша білуі.

Тәрбие басы – тіл. Алайда тілді тек коммуникативтік құрал ғана деп, бір жақтан түсіну тым жеткіліксіз. Олай дейтініміз, тілде екі түрлі білімдер жүйесі бар. Біріншісі – тілдік білім, яғни тілдің лексика-фразеологиясындағы мағыналар. Сөз мағынасы этномәдени тілдік ұжымның санасындағы ғасырлар бойы қалыптасқан білімнің айрықша түрі. Екіншісі – тілдік емес, дәстүрлі

білімдер жүйесі. Білімнің бұл түрі – этностын (халық, ұлт) өмір тәжірбиесінен ғасырлар бойы жинақталған, ұрпақтан-ұрпаққа мирас болып келе жатқан этнос санасындағы дәстүрлі дүниетанымы. Білім тілдік емес деп аталғанымен, ол ұжымдық санада тіл арқылы орнығып, тіл бірліктері арқылы объективтенеді. Лингвомәдениеттанушы бұл жерде тілді емес, тілдік материал арқылы этнос санасындағы білімдерді анықтап, интерпретациялайды.

Жалаң тілдік білім жеткіліксіз. Яғни лексика, фразеологиялық бірліктердің (единицалардың) мағынасын білу, яғни тілдік семантиканы білу жеткіліксіз, бұл ретте сол тілдік бірліктердің аясында құрылымданып, танымдық бірлікке айналған білімдердің – мәденионимдер семантикасын айқындаудың орны айрықша. Тек қана тілдік семантикамен мен шектелген тіл иелменінің мәдени кеңістігі, мәдени-тілдік құзіреті шектеулі болады. Екі түрлі білім – бірі тілдік (лексика-фразеологиялық мағына), екіншісі – тілдік емес дәстүрлі білімдердің бірлігі, бір-бірімен астаса келіп, санада мәдени-тілдік кеңістікті қалыптастырады.

Этнос санасындағы дәстүрлі білімдер жүйесі тілде, оның лексика-фразеологиялық бірліктерінде сақталған. Әртүрлі ұғымдар, атаулар арқылы санада тіркеліп, орнығып отырады. Ұғымдар түріндегі таным-білім сөз арқылы объективтенеді (репрезенттеледі). Бұндай ой-санадағы ұғымдар тілде лексикалық мағына түрінде қалыптасады да, тілдің лексикалық семантикалық бірлігіне айналады. Ал осы аталған лексика-семантикалық бірліктің аясында мәдени-ұлттық семантикалық құрылым пайда болады.

Лексикалық семантика мен мәдени семантиканы білдіретін таныммен байланысты тілдің лексика-фразеологиялық бірліктері ой-санада құрылымдана келіп, тілдік ұжымның санасында концептілік бірлік түзеді. Белгілі бір тілдік ұжымның санасындағы құрылымданған білімдер концепті формасында өмір сүреді. Әр халықтың, ұлттың тарихи даму барысында жасалып, тілдік санасында сақтаған мәдени-рухани байлығы болады. Заттық мәдениет, рухани мәдениет, дене мәдениеті дегендер – ұлттық мәдениеттің түрлері мен салалары. Этномәдени қордың дүниеліктері тілдің жинақтаушы, ұластырушы (трансформациялаушы) және қайта жаңғыртушы (ретрансляциялаушы) қызметі арқылы ұрпақтан-ұрпаққа мирас болып, ұлттың әрі қарайғы мәдени-рухани дамуының іргетасы болады. Олардың практикада, өмір тәжірбиесінде қолданылмайтын түрлері де жиі кездесуі мүмкін. Алайда ондайларды «күні өткен» керексіз дүниеліктер деп, материалдық және рухани құндылықтар қатарынан шығарып, мансұқ ету көрсоқырлық болар еді. Ондай архаикалық дүниеліктер ұлт мәдениетінің дамудың қандай сатыларынан өткенін, ұлттың өзін-өзі қалай дамытып отырғанын келер ұрпақ аңғара отырып, олар да өз кезегінде өзін-өзі танудың, өзін-өзі дамытудың жолын іздестіреді.

Халықтың дәстүрлі рухани мәдениетімен байланысты санадағы концепт түріндегі білімнің, жоғарыда айтылғандай, мәдени-тілдік бірліктер арқылы айқындалатыны белгілі. Алайда концептінің тілдік және тілдік емес мән-мағыналарын құрайтын мәдени код үнемі «ашық» бола бермейді, яғни өзінен-өзі түсінікті бола қоймайды. Бұндай мәдени семантикадағы жабық кодтарды ашу этномәдени концептілерді дұрыс интерпретациялауда өзекті мәселелердің бірі болып табылады.

Мәдени-тілдік таңбаларға мәдени нышанды лексика, фразеологиялық бірліктер, мақал-мәтелдер, нақыл сөздер (паремиологизмдер), тұрақты теңеулер, халықтық метафоралар, символдар, эталондар, прецедентті мәтіндер, прецедентті онимдер жатады. Ал халық шығармашылығы, фольклор, көркем сөз зергерлерінің көркем туындылары, айрықша эстетикалық таңбалар жүйесіне, тарихи, діни мифологиялық мәтіндер танымдық-тағылымдық таңбалар жүйесіне жатады. Осы таңбалардағы кодқа салынған мәдени ақпараттарды аша білу тіл иелменінің лингвомәдени құзыреттілігін көрсетеді. Бұны айтып отырған себебіміз, мысалы, қазақ халқының дәстүрлі мәдениетінде құдаласу, қалыңмал тәрізді ғұрыптық салттардағы мәдени кодтар дұрыс ашылмағандықтан, «құдалық», «қалыңмал» ғұрпы «қызды сату», «қызды саудалау» деп мәні теріс түсіндірілгені (интерпетацияланғаны) белгілі.

Ұлттық құндылықтар дүниесін аксиологиялық тұрғыдан тануда белгілі бір концептідегі мәдени-тілдік кодтарды ғылыми негізде ашудың тәсілдері мен жолдарын іздестіру, қарастыру, түптеп келгенде, мәдени-тілдік құзыреттілікке саяды. Ал мәдени-тілдік құзыреттілік жетпеген жағдайда этномәдени концептілердің мазмұны дұрыс ашылмай, теріс қорытынды жасалуы ықтимал. Мәселен, отырықшылық өмір жағдайымен байланысты, әсіресе, соқамен жер жырту, колөнер түрлері мен сауданың дамуы негізінде адамзат қоғамы әрі қарайғы прогреске қол жеткізсе, ал көшпенді малшылық артта қалып, тоқырауға ұшырап, көшпенді елдің мәдениеті мен тұрмысының дамуына кедергі болған деген біржақты тұжырым зерттеушілердің еңбегінде аз кездеспейді. Қазақтың көшпелі малшылық мәдениетіне «саяси-идеологиялық тапсырысты» айтпағанның өзінде, қайсыбір зерттеушілер аксиологиялық тұрғыдан дұрыс баға бермеді, себебі олар үшін мәдени код жабық еді. Мысалы, олардың ойынша, қазақ малшылары мал басын көбейте берген, яғни экстенсивті малшылық болған. Малшылықтың интенсивті түрін дамытпаған. Расында, қазақ малшылығында жабы көп өсіріледі. Өйткені қыр отына жайылған, бірнеше шақырымдық жайылымдағы шөптің алуан түрін шалып жейтін жылқының еті, қазы-қартасы аса дәмді, ал қымызы астың асылы болды. Қыр шөбіне жайылған жылқының азықтық сапасы отырықшылық мәдениетінде қолға қараған қолбала жылқының азықтық сапасынан әлдеқайда жоғары болған.

Сонымен қатар көшпенді малшылықты отырықшы егіншілікке қарама-қарсы қойып, көшпенді өмір салтын ұстанған елде мәдениет болмайды; рухани мәдениеті, саяси, қоғамдық көзқарасы төмен болады деп қорытынды жасайтын тұжырымдар тарихнамалық зерттеулерде аз кездеспейді. Сондай-ақ көшпелі қазақ қоғамының артта қалушылығы көшпелі малшылардың өндіріс күшінің, яғни еңбек құралдары түрлерінің жарамсыз және тым қарабайыр болуымен байланысты дейтін пікірлер де аз ұшыраспайды.

Ал бұндай сыңаржақ қорытындыдан Н.Н.Мацкевич, Г.Н.Семенюк сынды тарихшылардың пайымдауы мүлде басқаша. Мәселен, Г.Н.Семенюк өндіргіш күш тек еңбек құралдары ғана емес, адамның өзі де өндіргіш күшке жататынын, ал бұл жағынан көшпенділердің тәжірибесі, білімі, өндіріс дағдысы айтарлықтай жоғары деңгейде болғанын, осылар еңбек құралдарының орнын толтыратынын, тіпті көшпенді малшылыққа тән қарапайым еңбек құралда-

рының өзі өндіріс саласындағы өзіндік маңызын жоғалтпағанын атап айтады. Н.Н.Мацкевич, Г.Н.Семенюк қателескен жоқ. 50-70 шақырымдық ат жарыста, әскери жорықтарда қазақ малшысы өсірген қазақтың қыр аты қуса, жететін, қашса, құтқаратын жүйріктігін, күштілігін, ұзақ жолға төзімділігін көзбен көріп, көңілі сенген еуропалықтар оны жоғары бағалаған. Бұл – малшының дене еңбегі ғана емес. Бұл – олардың өндіргіш күші болып табылатын, Г.Н.Семенюк айтқандай, тәжірибесі, кәсіби білімі.

Бұл жерде ғалымның тәжірибесі, дағды, білім деп отырғаны – көшпелі малшы қауымның ежелден келе жатқан дәстүрлі білімдер жүйесі. Ол әралуандылығымен ерекшеленеді: халықтық ветеринария, халықтық зоотехника (малдан алынатын өнім түрлерін тиімді тәсілдерімен өндіру), қырдағы-ойдағы нәбағаттар мен хайуанаттар әлемі, халықтық астрономия, халықтық метеорология; жер бедерін аудандастыратын халықтық география тәрізді малшы қауымның жыл он екі ай табиғатпен бетпе-бет келуі, өмір тәжірибесі дәстүрлі білімдер жүйесін қалыптастырған.

Малшы қауым төрт түліктің өзді-өзіне тән ерекшеліктері, жейтін отының түр-түрлері, жыл бойы жейтін шөбінің түрлері мен маусымдық шөбі; жайылымның маусымдық түрлері (көктеу, жайлау, күздеу, қыстау), шаруашылық жүргізудің маусымдық іс-шаралары (қысырақ матау; үйір құру; айғырды үйірге салу, қосқа шығу және т.б), өріс аудару, жайылым қуалу, күйек байлау, күйек шешу, күйек алу және т.б. тәрізді маусымдық науқан жұмыстары өмір тәжірибесінде сыннан өткен білімді қажет етеді. Сондай-ақ бұндай аса маңызды іс-шаралар мен жұмыстарды шаруашылық жүргізудің арнайы күнтізбесінсіз атқару мүмкін емес. Шаруашылық күнтізбесіне сүйенбеген жылқышы құрығын, қойшы таяғын ұстап, малынан айырылып қалады. Сондықтан көшпенді малшы қауым қазақ сахарасының табиғи ерекшелігіне байланысты құрылған тоғыс есебі деп аталатын арнайы күнтізбе жүйесін жасады. Халықтың астрономиялық білімдер жүйесі негіз болған тоғыс есебі, түптеп келгенде, өндіргіш күштің дамуына ықпал еткен әрі көшпенді малшылықтың рухани-мәдени дүниесінің бірі ғана емес, бірегейі болып табылатыны сөзсіз.

Қысқасы, осы айтылғандарға көз жеткізу үшін және аксиологиялық тұрғыдан объективті баға беру үшін Ұлттық корпуста «Тоғыс есебі», «Үркер» тәрізді ұғымдық, танымдық бірліктердің мәдени кодын ғылыми-теориялық жағынан төмендегіше талдауға болады:

«ҮРКЕР» концептісі. Үркер – күншығыстан таң біліне көкке шығып, күнбатыстан жерге түсетін, халықтың байырғы түсінігінде аса қадір тұтатын шоғыр жұлдыздардың бірі. Қазақ есепшілері қырық күн шілденің басталғанын ҮРКЕР-ге қарап біледі. Өйткені, мамырдың 10-12-лерінде ҮРКЕР жерге түседі де, содан 40 күн жерде жатып, маусымның 20-22-лерінде таң ата шығыс жақтан азғана уақыт көрінеді де таң жарығымен бүркемеленіп қалады. Аспан шырағының бұл көрінуін есепшілер Үркердің көкке шығуы дейді. Осы кез күннің ең ұзарған, түннің ең қысқарған мезгілі. Бұл кезде кірген 40 күн шілде тамыз айының 2-4-інде шығады. Сөйтіп, ҮРКЕР көк жүзінде маусымның аяғында шығыстан таң қараңғысында, түн ортасында, түн ортасына жетпей көрінеді. Қыркүйек айының соңына қарай түстік шығыстан қас қарая көрінеді

де, бір топ жұлдызды – Үш Арқар – Таразы, Сүмбілені бастап, батысқа шеру тартқан қыстық сапары басталады. Үркердің тоқ басы батысқа қарап тұрады.

Көк жүзінде ҮРКЕР бастап қиғаштай туған Үш Арқар – Таразы, Сүмбілені көрген есепшілер жұлдыздар тіркесіпті дейді. Көк жүзіне көз жіберген адам қырдан қиғаштай асып бара жатқан түйелі көштің елесін көргендей болады. Осы кезде ҮРКЕР-ді қара өлеңге қосқан өлеңшілер:

Көп жұлдыздың ағасы Үркер деген,
Түйені сәнді көшке тіркер деген.
Сұлуды аш белінен ұстап тұрып,
Бетіне бес қалампыр бүркер деген, –

дейді. Осылайша халық тілінде «Үркер – жұлдыз ағасы» аталып кеткен. Содан бірінің соңынан бірі тіркескен түйелі көшке ұқсаған жұлдыздар наурыз айының аяғына таман батыс көкжиекте бір түзудің бойына келіп қатар түзейді. Есепшілер бұл көріністі «жұлдыздар қатар түзепті», «жұлдыздар теңесіпті», «күн мен түн теңескен» екен дейді. Есепшілер күн мен түннің теңелгенін осы үш жұлдыздың қатар түзегенінен байқайды.

Сонымен, қыркүйектің аяғында бір топ жұлдызды бастаған ҮРКЕР-дің батысқа тартқан шеруі мамырдың 10-12-лерінде аяқталып, ҮРКЕР жерге түседі. Бұл – тоғыз есебі бойынша ескі жылдың аяқталғанын, сондай-ақ, В.В.Радлов, М.Ысқақов, Х.Әбішев сынды ғалымдардың деректерінде айтылғандай, Жаңа жылдың басталғанын білдіреді.

Есепшілер ҮРКЕР-дің шығыстан батысқа тартқан «қысқы шеруін» Үркерлі айдың бәрі қыс деп, малшы, егінші қауымды сақтандырып отырады. ҮРКЕР төбеден ауып, батысқа қарай жамбастаса, байланған күйек алынады дейді. Яғни ҮРКЕР көтеріліп, қошқар тұмсығын жоғары көтергенде маусым шыға байланған күйек 90 күн күн шамасы өткен соң, қазан шыға ҮРКЕР шығыстан көз байлана көрінген кезде шешіледі. Бұл кезде қошқардан шыққан (қашқан) қой наурыздың аяғына таман жаппай қоздай бастайды. Бұл шақты күйек алу, ал қозыны күйек қозы дейді.

Үркер жерге түспей жер қызбайды; Үркер жерге түскенде қой қырқылады; Үркер туса сорпа ас болады; Үркер туса қозы піседі; Үркер көтерілгенде шөп те көтеріледі; Үркер көтерілгенде бидай бас тартады, Үркер көтерілгенде қошқар басын көтереді, яғни «қошқарға күйек байланатын кез болады» деген мағынаны білдіреді. Осылайша есепшілер шаруашылық ұйымдастыруда уақыт жағынан ҮРКЕР-ді межеге алып, оның тууын, көтерілуін шығыстан, батыстан, қақ төбеден көрінуін бағдар тұтқан. ҮРКЕР қас қарая күзде (қазан айының басында) шығыстан, қыстың ортасында төбеден, көктемнің басында батыстан көрінеді. ҮРКЕР-дің осы көрінуін бағдарға алып жыл мезгілдерін анықтайды. Үркер – жыл мезгілдерін жазбай тануға өте қолайлы жұлдыз.

Халықтың паремиологиялық қорына қазақ есепшілерінің кіргізген бұндай тұрақты сөз орамдары мен мақал-мәтелдері аз емес. Сондай-ақ ел ішінде тәуліктің түнгі шақтарын жолдағы түн қатқан жолаушы, тебінде жүрген жылқышы ҮРКЕР арқылы да анықтап отырған. Мысалы, тамыз айының

кірісінде түн ортасына жете көрінген туған (көрінген) ҮРКЕР батыс көкжиекке тақап батқанша (көрінбей кеткенше) ұзамай таң да атып қалады, тәуліктің бұл шағын таңғы бір ұйқы, жарым ұйқы қалды дейді.

Желтоқсанның кірісінде қас қарая шығыстан ауа көрінген ҮРКЕР таң ата батады. Бұл кез – Үркердің түнгі аспандағы ең ұзақ жолы, басқаша айтқанда, түннің ең ұзақ шағы басталған шама. Әсіресе, қостағы жылқышы таңның атуын ҮРКЕР-ге қарап сарыла күтеді. Қаңтардың ішінде іңір қараңғысында төбеден көрінген ҮРКЕР таңғы үш ұйқы қалғанда (түнгі сағат үштер шамасында), таңға жетпей батып кетеді.

Халық астрономдары (есепшілер) ҮРКЕР мен Айдың тоғысын уақыт өлшемінің эталоны етіп алып, тоғыс есебі деп аталатын шаруашылық күнтізбесін (календарын) жасаған.

ҮРКЕР-мен байланысты айтылған жайттар ертегі, жұмбақ, айтыс, кара өлең тәрізді халық шығармашылығында жиі ұшырасады. Өлеңшілер жұмбақ өлеңде жаздың қырық күнінде көк жүзінен көрінбей кеткен ҮРКЕР-ді жұмбаққа арқау еткен:

Қысырақ жоғалды жаз, кетіп жырақ,
Көп жүрдім, іздеп оны елден сұрап.
Тұр екен дарияның ортасында,
Жете алмай, үйге қайттым босқа бірақ.

Ал жұмбақ шешуші:

Үйірі қысырықтан Үркер деймін,
Мен де оны жазда көктен бір көрмеймін.
Болады ол алты ай көкте, алты ай суда
Өлеңім осылай жөн сілтер деймін, –

деп шешуін тап басып таба қояды.

Халықтық этимологияда ҮРКЕР-ді «малдың үркіп бөлініп қалған үйірі» деп «төркіндетеді». Осы түсінікпен байланысты болса керек, жұмбақшылар ҮРКЕР-ді қалың жылқыдан бөлініп қалған қысырақтың үйірі деп жұмбақтаған.

Аңыз әңгімелерде ҮРКЕР қырық күн жерде жатып, шөпті қаулатып айдап жер бетіне шығарады екен; он екі жылдағы хайуандар ҮРКЕР-ді жерден шығармай басып тұрады екен, алайда кезек сиырға келгенде, аша тұяқты болғандықтан, басып тұра алмай, көк шығарып алыпты, ҮРКЕР-лі айдың бәрі қыс болатыны сиырдың кесірінен екені; ҮРКЕР әуелде жетеу екен, одан бөлініп шыққан Сұлусарыны, – кейде Үркердің үлпек сарысы деп те айтылады, – Жетіқарақшы ұрламақ болып қуып жүреді екен делінеді.

Ай мен ҮРКЕР өлең-жырларда көбіне қатар айтылады. Әдетте Ай ҮРКЕР-дің оң жақ тұсынан маңайлап келеді. Көк жүзіндегі бұл көріністі Ай ҮРКЕР-ге «ауыл үй қонды» дейді. Айдың көк жүзінде ҮРКЕР-мен бетпе-бет келетін сәті болады. Бұл көріністі Ай мен ҮРКЕР «қойындасты» дейді. ҮРКЕР-мен қойындасқан Ай ертеңіне-ақ одан «тұра қашқандай» болып жырақтап кетеді. Осы көріністі өлеңшілер:

Жүруші едік балдай боп,
Құлын менен тайдай боп,
Арамыз кетті қашықтап,
Үркерден қашқан Айдай боп, –

деп қыз бен жігіттің арасындағы байланысты ҮРКЕР-ден жырақтаған Айға теңейді. Ал Ай мен ҮРКЕР қойындаспай жақын маңайласа, ертеңіне жырақтап кетпей, бір-біріне жақындау тұрады.

Қазақ ауыз әдебиеті мұраларын молынан жинаған этнограф ғалым Ә.Диваевтың қолжазбаларындағы «Алыптар туралы» немесе «Кім күшті?» деп аталатын ертегіні академик Ақжан Машанов (1906-1997) халықтың астрономиялық жұмбақ ертегісі дей келіп, оны бала кезінде естігенін де, өзі де бірталай адамға айтып бергенін еске ала отырып, аталмыш ертегіні қысқаша былай деп мазмұндайды:

«Ерте заманда бір кісінің үш баласы болады, бір алып көк өгізі болады. Өгіздің алыптығы сонша, оның басы мен құйрығының арасы бір күншілік жол болса керек. Өгізді ағайынды үш жігіт бағады: басын – үлкені, кіндік тұсын – ортаншысы, бөксе жағын – кенжесі. Күндердің күнінде кенже інісі бір түс жүріп, ортаншы ағасына келеді де, өгіздің жапаламай тұрғанын хабарлайды. Екеуі қосылып, үлкен ағасына келіп, өгіздің жапаламай қойғанын және бүйірі ортайып, ашыққанын хабарлайды. Ол өгіздің жайылмай тұрғанын айтады. Үшеуі жетелеп, өгізді суаруға бір көлге апарды. Өгіз көлдің суын тауыса ішіп қойғанда, оның түбінен бір жайын шыға келіп, өгізді жұтып жібереді. Көл таусылған соң оның маңындағы құстар да ұшып кетеді, *el de* тарайды.

Сол екі ортада күннің көзін қара бұлттай қаптап, қанатынан қар жауған, тұмсығынан мұз тамған керемет бір алып қаракұс келіп, өгізді жұтқан жайынды іліп алып, көкке шырқай жөнеледі.

Бір таудың бетіне жайылған бір қора қой. Таудың қақ басында бір алып көк серке тұр. *Сол* серкенің бауырында жел-жауыннан ықтаған қойшы отыр. Жайынды ілген алып қара құс сол серкенің мүйізіне келіп қонады да, жайынды да, оның ішіндегі өгізді де боршалап жей бастайды. *Сол* кезде өгіздің жауырын сүйегі келіп қойшының кезіне түсіп кетеді. Қойшы (бай) түрегеліп, қойын айдап үйіне келеді де: «Көзіме қықым түсті» – дейді. Әйелі (бәйбіше) оның көзіндегі жауырынды алып тастайды. Ауыл ол жерден көшеді, жауырын жұртта қалады.

Ауылдың жаңа қонған қонысы айналасы қырық күншілік көлдің жағасы болады. Бәйбіше қырық күншілік көлден көлденең аттап өтіп, қырық қап тезек тереді.

Ескі жұртта қалған баяғы жауырынның үстіне түн қатқан қырық жолаушы қонады. Таң ата бастағанда жолаушылар жер сілкініуден оянады. Қараса, бұлардың тақыр жер деп қонғаны алып жануардың жауырын сүйегі екен. Соны бір күшік түлкі кеміріпті – жер сілкінгені сол екен. Бәрі жабылып түлкіні соғып алады. Түлкінің үстіндегі терісін сойып, қырық жолаушы бөліседі, аттары көтергенше артып алады (кейбіреулердің айтуынша, әрқайсысы бірден қырық ішік, қырық тымақ шығатындай тері алады). Түлкінің астыңғы терісі сойылмай қала береді. *Оны* қойшы бай тауып, сойып алады (кейде бәйбіше тауып алады).

Жолаушылар жоғарыда айтылған қойшы бай мен бәйбішенің аулына келеді. Келсе, бәйбіше босанып, баласы қырқынан шығып, соның тойы болып жатыр екен. Түлкінің жарты терісін балаға бөрік істемекші болып, басына өлшегенде, маңдайына ғана шығады, *арты* жетпейді. Жолаушылар өздері алған теріні бәйбішеге сыйлайды. *Сонда* ғана түлкінің тұтас терісі балаға жұрын бөрік шығады».

Ертегі осымен аяқталып, қорытындысында: Кім күшті? Осы алыптардың қайсысы зор деген сұрақ қойылады.

Осы айтылған ертегідегі алыптарды аспан кеңістігіндегі шокжұлдыздармен байланысты деп тереңнен пайымдай отырып, астрономиялық ғажайып аңыздың сырын ғалым асқан білгірлікпен былай деп түсіндіреді:

Көк өгіз көкке қарап қаңтарылады [Көк өгіз деп Үркерді айтып отыр – ред.] Қаңтардың айы туды, – қыс басталды, өгіз ашықты. Көл суалды, су қатты. Жайынның өгізді жұтуы – сол кезде ҮРКЕР-дің жайын жұлдыздар тобына деңгейлесуі немесе тоғысуы. Бұл – ноябрь, декабрь айының кезі. Жайынды және онымен қабат өгізді қара құс көтеріп әкетуі ҮРКЕР-дің онан да жоғары шырқағанын көрсетеді. Оның көк серкенің мүйізіне барып қонуы – қыс ортасында ҮРКЕР-дің тас төбеге барып тоқтаған кезі, күннің ең қысқарып, түннің ең ұзарған кезі. Бұл – жаңа жыл мейрамы (декабрьдің аяғы, январьдің басы), Таутеке (Козерог) айы [.....]. Жылға есептегенде, ол – Қаракиік (Доңыз), жұт болатын жыл [.....]. Қыстан қысылған қойшы серкенің төрт аяғының астына тығылады. Дәл қыс ортасы кезінде Қарақұс та (Орел), Көк өгіз де, Жайын да Көксеркенің мүйізі үстіне туралап келеді.

Көк өгіздің жауырыны келіп, қойшының көзіне түскені – ол Үркердің жамбасқа түсіп, көз деңгейіне келгені болар. Малдың жауырыны көрінуі – ҮРКЕР-дің көлденеңдеп, жамбасқа түсуі (басы емес, аяғы емес, құйрығы емес, жауырыны көрінеді). Жауырын сүйек – малдың иесіндей, қасиетті сүйек саналады. Қойшы оны ояу отырғанда көрді (көзі жұмулы болса, жауырын түспес еді). Демек, «Жатқанша Үркер жамбасқа келсе, жаз шықпағанда не қалды?» – деп, қойшы қозғалды да, көк серкенің бауырынан шығып, жаз қамына кірісті. Қыстан қысылған бақташы жаз шыққанға жадырап, етекке қарай құлады, көшті. Шаттық жырын жырлады:

Қойшы қойын келеді таудан жайып,
Көк серкесі қалыпты мұзға тайып.

Осы өленді жас кезімізде қарт адамдардан талай естіген едік. Көшкен ел айналасы қырық күншілік көлге қонды. Көл – көктің көлі. *ҮРКЕР*-ді халық суға түседі деседі. Айналасы 40 күндік дегені – 40 күн шілде, ол ҮРКЕР-дің көрінбейтін кезі. Ақ молайды: «сүт көп, көмір аз», құрт қайнатуға отын керек, сондықтан бәйбіше күнде тезек тереді. Өгіз жауырынының үстіне қонған 40 жолаушы да сол 40 күн шілдені көрсетеді.

Жауырынның қайта қозғалуы, жолаушылардың оның үстінен кетуі шілденің аяқталып, *ҮРКЕР*-дің қайта көріне бастауы. Жауырынға қозғау салуы – кеміруші күшік түлкі болу себебі дәл сол шілде өткен кезде Күшік түлкі дейтін жұлдыздар тобы төбеде болады.

Түлкінің екі жақ терісі баланың басына өлшенуі – күздігүні күн мен түннің теңелген кезін, Таразы айының туғанын көрсетеді. Ерте кезден белгілі, адамның беті күндізге және жазға, желкесі түн мен қысқа ишарат.

Түлкінің бір жақ терісін баланың басына өлшегенде, оның маңдайы мен екі құлағы шығып, *арты* жетпей қалады. Желке – түн белгісі, оның жетпей тұруы түннің қысқарып, күннің ұзарғанын көрсетеді. 40 жолаушының алған үлес терісін құрағанда ғана, алды-*арты* теңгеріліп, бөрік шығады, яғни шілде өткесін, түн біртіндеп ұзарып, Таразы айында күн мен түн теңгеріледі.

Бұл тұста көрнекті ғалымның этномәдени интерпретациясына алып қосарымыз жоқ. Әйтсе де, бәрінен де ҮРКЕР-дің ерекшеленетінін атап айтқымыз келеді. Халықтың даналығын, есепшінің (халық астрономының) ұшқыр да тапқыр қиялын тереңнен танытатын теңдесі жоқ астрономиялық ертегінің негізгі түп қазығы – ҮРКЕР. Оның жайынға жұтылуы, көкке көтерілуі (Жайын шокжұлдызымен маңайласуы), одан көк серкенің мүйізіне шығуы. Одан жауырынға айналып, қиқым болып көзге түсуі (жамбасқа келуі), оны көзден алып тастау (жерге түсуі), жерде қырық күн жатуы, оның қайта қозғалуы (көкке шығуы), ҮРКЕР-дің жылдық қозғалысын өзге де шок жұлдыздармен маңайластығын аңғартатын мифтік формадағы ерекше көркем «сюжеттер». Сондай-ақ түннің ұзаруы, күннің қысқаруы, керісінше, күннің ұзарып, түннің қысқаруы; есепшілер метеорологиясы танытатын басқа да жайттар ҮРКЕР-дің көк жүзіндегі сапарымен орайластырылған.

Міне, халықтың шығармашылық дүниесінде ҮРКЕР-дің ерекше орын алуы аталмыш аспан шырағының ел өмірінде аса маңызды болғанын байқатады. Атап айтқанда:

– халық астрономдары – есепшілер ешбір баламасы жоқ тоғыс есебі деп аталатын шаруашылық календарының (күнтізбесінің) жүйесін жасаған;

– Үркердің батуына, тууына, көрінуіне қарап, шаруа адамы маусымдық шаруашылық жұмыстарын ұйымдастырған;

– жыл мезгілдерін анықтауда Үркерді эталон етіп алған. Кейбір халықтар жыл мезгілдерін *көк шықты, құс келді, құс қайтты, қар түсті* деп, жердегі экологиялық (табиғи) өзгерістерге сүйенсе, қазақ есепшілері көк жүзіндегі жұлдызды, яғни Үркерді бағдар етіп алған;

– түнгі уақыттың қай шама, қай шақ екенін Үркерге қарап анықтап отырған. «ТОҒЫС ЕСЕБІ» концептісі. Тоғыс есебі – байырғы қазақ қоғамында Үркер жұлдызы мен Айдың тоғысуына негізделген халықтық күнтізбе. Ай өзінің әр кезеңінде (фазасында) Үркер жұлдызын басып өтеді, яғни есепшілердің тілімен айтқанда, Үркер мен Ай тоғысады. Алайда Үркердің көк жүзінен көрінбей кететін кезі болады. Мамырдың (майдың) 10-12-лерінде батыс көкжиектен төмендеп, кеш қараңғысында көрінбей кетеді. Содан Үркер 40 күн бойы, маусымның 22-сіне дейін Жерде жатады. 22-24 шілдеде (июльде) таң ата Шығыстан Үркер көкке шығады. Осы кезден бастап Үркердің көк жүзін кезген ұзақ «сапары» басталады. Осы сапарында, Тоғыс есебі бойынша, Үркер мен Ай көк жүзінде он бір мәрте тоғысады. Үркер әуелде көк жүзінде жиырма бірдегі Аймен, яғни Ай табағының жиырма бірінші кезінде (фазасында) тоғысады. Бұл кез Ай табағының оң жақ (батыс жақ) шетінен «таусылып, мүжіле» бастаған кезі.

Ай өзінің жиырма бір жаңасында шілде айының ішінде Үркермен тоғысады. Тоғыстың бұл айын жиырма бір тоғыс деп атайды, қазіргіше шілде айына сәйкес келеді. Тоғыстың бұл айында шілде шығады, шілденің жиырмаларында Таразы (Нарық жұлдызы) туып, таң суиды.

Айдың Үркермен келесі тоғысы Айдың он тоғыз жаңасында өтеді. Айдың Үркермен кездесуі (тоғысуы) жай адамға бұл кезде анық, айқын көрінбейді, оны есепшілердің жіті көзі ғана байқайды. Тоғыстың бұл айы он тоғыз тоғыс деп аталады, қазіргі тамыз айына сәйкес келеді. Тамыздың жиырмаларында Сүмбіле (Теке) туып, су суиды.

Айдың Үркермен келесі ұшырасуы Айдың он жеті жаңасында өтеді. Тоғыстың бұл айы он жеті тоғыс деп аталады, қазіргіше қыркүйек айына сәйкес келеді. Тоғыстың бұл айында қазанның жапырақ желі, күземшуақ, күзгі сұлбі, боз қыраудың амалы, киіктің матауы тәрізді амалдар өтеді.

Айдың Үркермен келесі кездесуі Айдың он бес жаңасында өтеді, яғни Үркер толған Аймен ұшырасады. Тоғыстың бұл айы он бес тоғыс деп аталады, қазіргіше тоғыс қазан айының ішінде өтеді. Бұғының мойын жуары, боқырау амалдары болады.

Үркердің кезекті кездесуі он үштегі Аймен, толықсыған Аймен, Айдың ең сұлу кезінде өтеді. Тоғыстың бұл айы он үш тоғыс деп аталады, бұл тоғыс қазіргіше қараша айының ішінде болады. Қарашаның ала шабыр қары, теке бұрқылдақ амалдары өтіп, жолдас қар түседі.

Айдың Үркермен келесі ұшырасуы Айдың он бір жаңасында өтеді. Тоғыстың бұл айы он бір тоғыс деп аталады, қазіргіше тоғыс желтоқсан айының ішінде өтеді. Қаңтардың қарлы бораны, күннің қаңтарылуы (тоқырауын) осы тоғыста болады.

Айдың Үркермен келесі ұшырасуы Айдың тоғыз жаңасында өтеді. Тоғыстың бұл айы тоғыз тоғыс деп аталады, қазіргіше тоғыс қаңтар айының ішінде өтеді. Тоғыстың бұл айында өтетін амал – ай мүйізді алты күн, ақпан-тоқпанның аязы. Бұл кезде Үркер төбеде тұрады.

Айдың Үркермен келесі ұшырасуы Айдың жеті жаңасында өтеді. Тоғыстың бұл айы жеті тоғыс деп аталады, қазіргіше тоғыс ақпан айының ішінде өтеді. Бұл тоғыста қайталаманың ақ бораны болады дейді.

Айдың Үркермен келесі кездесуі Айдың бес жаңасында өтеді. Тоғыстың бұл айы бес тоғыс деп аталады, тоғыс наурыз айының ішінде болады. Үркер төбеден ауып, бес тоғыстың бет жуары, отамалы ойылған қыс амалдары өтеді, итшуақ, кемпіршуақ сияқты шуақты күндер болады, қарды әкететін алтынкүрек желі еседі. Ескек соғады.

Ай өзінің үш жаңасында Үркермен кездеседі. Тоғыстың бұл айы үш тоғыс деп аталады, Тоғыс, қазіргіше сәуір айының ішінде өтеді. Үркер жамбасқа келеді, құралайдың салқыны, құс қанаты, сары шымшықтың салқыны тәрізді амалдары болады деседі.

Ай өзінің бір жаңасында батыс көкжиекте жерге батып бара жатқан Үркермен ұшырасады. Тоғыстың бұл айы бір тоғыс деп аталады, тоғыс қазіргіше мамыр айының ішінде өтеді. Үркер жерге түседі, Үркер суға батса, күн жауыншашынды болады, жерге батса, күн құрғақ болады дейді.

Сонымен, Үркердің түстік Шығыстан басталып, батысқа шеру тартқан сапары аяқталып, Үркер жерге батады. Оңтүстік Шығыстан кеш қараңғысында көрінген Үркердің түстік батыстан жерге батқанша көк жүзінде көрінуін есепшілер Үркерлі айдың бәрі қыс деп, сақтандырып отырған. Үркер жерге түскенде ғана жаз шықты деп, ел көктеуден жайлауға көшетін болған.

Тоғыс есебі, жоғарыда көрсетілгендей, кемімелі тақ санмен есептеледі. Есепшілер тілімен айтқанда, Үркер таққа тоғаяды.

Үркер тоғысының, әсіресе, малшы қауымы үшін біраз артықшылығы байқалады. Жыл басының Үркердің батуының көктемнің соңғы айына келуі ел тұрмысы үшін қолайлы болуымен бірге, есепшілер ендігі жерде Үркердің тәуліктік қозғалысын астрономиялық меже етіп алып, түннің ұзақтығы, қысқалығы тәрізді құбылыстарды анықтаған. Бұрын түннің ұзақ не қысқа екендігін жай ғана сезінсе, ендігі жерде бұл құбылысты Үркер арқылы белгілеу мүмкін болды. Үркер қас қарая көрініп, таң атқанша батпайтын болса, ең ұзақ түн (қыс ортасы), ал түн ортасы ауа батып кетсе, күннің ұзарып, түннің қысқара бастағаны. Сөйтіп, есепшілер Үркердің жылдық жолына ғана емес (мысалы, Үркер жерге түскен кезден келесі жылғы Үркер көкке шыққан кезге дейінгі аралық), тәуліктік қозғалысына да назар аударады.

Үркер тоғысының тағы бір артықшылығы – ауа райын болжауға, жыл мезгілдерінің (солярлық уақытты) қалай өтетінін жорамалдауға қолайлығы. Үркер есебі бойынша жыл екі мезгілге бөлінген. Үркер «жерге түсіп», көрінбей кеткен кез – жаз, ал Үркер көрінетін байланысты кез – қыс. «*Үркерлі айдың бәрі қыс*» деп айтылатын дағдылы сөз орамы осыны аңғартады. Ана тіліміздің асыл қорында Үркерге қалыптасқан сан алуан сөз тіркестері бар. Ондай сөз орамдарының жылдың қай мезгіліне, айына қатысты айтылатынын Тоғыс есебі-нен хабардар адам бірден аңғарады. «Толған ай тұрса, Үркерге қарсы карап, қыс түсер шақырайын боран сабап» деген сөз тізбегі он үш тоғыс айын (ноябрь) білдіреді. «Үркер үйден көрінсе, үш ай тоқсан қысың бар» дейді. Бұл кез – тоғыз тоғыстың кезі, өйткені қаңтар айында Үркер қақ төбеде тұрады.

Сескенер өлім десе шыбындай жан.

Сарғайып Үркер ауса атады таң, –

дейді Мұрат Мөңкеұлы. Майталман ақынның жылдың қай мезгілін, қай айын айтып отырғаны қазіргі оқырманға жұмбақ тәрізді көрінуі ықтимал. Ал, шындығында, Үркер ауа таң ататын мезгіл бес тоғыс (март) айына сәйкес келеді. «Үркер жамбасқа келсе, жаз шықпағанда несі бар» деген нақыл үш тоғыс (апрель) айына қатысты айтылған.

Есепшілер жыл сайын белгілі бір уақытта қайталанып отыратын ауа райындағы құбылыстарды (жауын-шашын, жел, боран, ыстық, суық т.б.) амал деп атап, оның қай кезде болатынын Тоғыс есебі бойынша анықтап отырған. Сөйтіп, олар күнқайыру жүйесіне амал деген жаңа бір төл ұғымды енгізді. Шаруашылық ұйымдастыруда, әсіресе малшы қауым Тоғыс есебіне сүйене отырып, әр айдағы амалдардың қалай өтетініне мұқият көңіл бөлген.

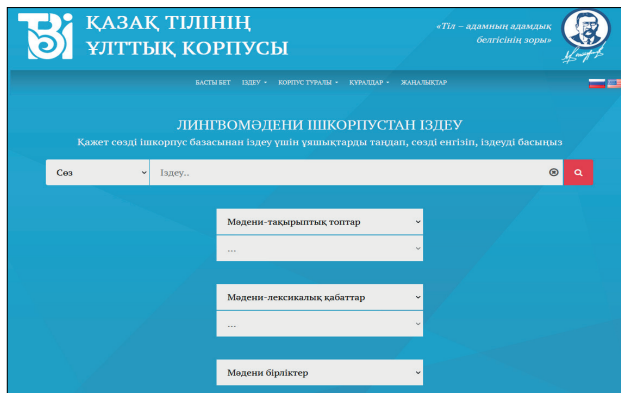
Қорыта келгенде, жоғарыда жабық кодта тұрған мәдени ақпараттардың жаһандану тұсында ақпараттық-цифрлық қорлардың бірі және бірегейі – Қазақ тілінің ұлттық корпусында көрініс табуы бүгінгі ұлттық сананы жаңғыртудың және ұрпақтан-ұрпаққа жеткізудің тиімді тетігі болмақ.

8.3 Лингвомәдени ішкорпустаың мәтіндік-лексикографиялық базасын әзірлеу

Ақпараттық технологиялар дамыған сайын адамның өмір сүру деңгейі ілгерілеп, кез келген қызмет түрі жеңілдеп, оңтайландырылды. Нәтижесінде қоғамға қажетті ғылым салаларының барлығын цифрландыру қажеттілігі туындады.

Көлемді материалдарды пайдалануға, қажетті тілдік деректерді тез тауып алуға және оларға тиісті өңдеулер жүргізуге мүмкіндік тудырып, зерттеу тәсілдеріне жаңаша көзқараспен қарауға және ғылыми айналым аясына аса маңызды тілдік материалдарды енгізуге жағдай жасау мақсатында профессор А.Жұбанов қазақ тілінің корпусын әзірлеу мәселесін көтерді. «Ұлттық тіл мәтіндерінің компьютерлік корпусын құру жобасы бір ғана ғылыми ұйымның шешетін мәселесі емес және ол зерттеу жұмысы 3-5 жылда аяқтала қояды деуге де болмайды. Себебі бұл аса күрделі және оның нәтижелері әлемдік дәрежедегі аса маңызды ғылыми жұмыс болып саналады. Зерттеу жұмысының мақсатына сай орындалатын міндеттері сала-салаға, кезең-кезеңге бөлініп, тек қана ғалымдар қауымдастығын құру арқылы ғана ауқымды нәтижеге жетуге болары сөзсіз. Қазақ тілінің ұлттық корпустарын жасау – көп болып жұмылып атқаратын ұлттық құндылығымыз. Сондықтан Қазақстанның әр жерінде бірінен-бірі дербес атқарылып жатқан корпус жасау ісін орталықтандыру керек. Оған Қазақстанның әр жерінде ғылыми-педагогикалық қызмет атқарып жүрген ғалымдардан арнайы лингвистикалық топ құру керек. Өйткені тілдік бірліктерді модельдеу – өте күрделі мәселе. Сондай-ақ тілдік корпустар құрастырудың өзі ең алдымен лингвистикалық белгіленім талдамасын жасауға тіреледі. Осы кезге дейін жасалып жатқан корпус жасау тәжірибесіндегі нәтижелерді бір орталыққа жинақтау керек. Кітап, газет-журнал шығаратын баспалармен шартқа отырып, мәтіндердің электронды нұсқаларын алу қажет» [12, 4] екенін айтқан.

Осыған байланысты А.Байтұрсынұлы атындағы Тіл білімі институтында тілтанымдық ІТ ресурс – «Қазақ тілінің ұлттық корпусы» әзірленіп, көпшілікке ұсынылып отыр. ҚТҰК-да қазақ тілінің 40 миллион сөзқолданыстан тұратын электронды мәтіндер қоры жинақталған (<https://qazcorpus.kz/>). Оған енгізілген мәтіндер қорына морфологиялық, лексикалық, семантикалық, фонетика-фонологиялық белгіленімдер қойылған. Сондай-ақ ҚТҰК-ның құрамына енетін негізгі корпустан қатар диалектілік, тарихи, ауызша, параллель, Ахмет Байтұрсынұлы, ономастикалық, мақал-мәтелдер, жарнама және лингвомәдени ішкорпустар да әзірленіп жатыр. Біздің мақаламыз ҚТҰК-ның (әрі қарай ҚТҰК) құрамына енетін *Мәдени-репрезентативті мәтіндер ішкорпусының* лексикографиялық базасын әзірлеудің ерекшеліктеріне арналып отыр.



Сурет 24. Лингвомәдени ішкорпустаң интерфейсі

Лингвомәдени ішкорпус әзірлеушілердің басты мақсаты – қоғамның ұлттық бірегейлігін нығайту, келешек жастардың рухани санасын жаңғырту, ұлттық рухты көтеріп, отаншылдыққа тәрбиелеу, лингвистикалық, мәдени-этнографиялық, тарихи-танымдық ғылыми зерттеулер санын бірнеше есеге арттыру, білім беру-оқыту үдерісін айтарлықтай жеделдету, мемлекеттік тілдің мәртебесін нығайту, сол арқылы әлемдік және отандық мәдени кеңістікте беделін көтеру болды. Әлемнің ешбір тілінің корпусында жоқ Лингвомәдени ішкорпуста әзірлеу – А.Байтұрсынұлы атындағы Тіл білімі институтының директоры А.Фазылжанның идеясының негізінде жүзеге асырылып отыр.

Этнос болмысын тану тіл арқылы жүзеге асатыны белгілі. Тіл арқылы сол ұлттың өзіне ғана тән ерекшелігі, осы уақытқа дейін айналысқан тұрмыс-тіршілігі, ұстанатын діні мен сенімі, құндылықтары, салты мен дәстүрі сыртқа шығады, өзгелерге жетеді. Ұлттық код та тілде сақталады. Белгілі ғалым Ә.Т. Қайдар: «Этностың басып өткен сан ғасырлық даму жолы, оның белгі-бейнелері біздерге тас мүсіндер мен жартастарға қашалған сына жазулар арқылы, мәдени ескерткіштер мен әртүрлі ғимараттар түрінде жетуі мүмкін. Бірақ бұлардың бәрі этнос өмірінің мың да бір елесі ғана. Оның шын мәнісіндегі даналығы мен дүниетанымы тек тілде ғана сақталатынын» [197] айтады. Тіл – тек қарым-қатынастың құралы емес, ол сол тілде сөйлейтін халықтың тарихы, болмыс-бітімі, салт-дәстүрі, табиғаты, ұлттық ерекшелігі, тұрмысы мен тіршілігі.

Ақиқат болмысты танып-түсіну нәтижесінде жинақталған ақпарат тілде көрініс табады. Ғаламның тілдік бейнесі негізінен субъективті сипатта, себебі онда нақты бір адамның көзқарасы, дүниетанымы көрінеді. Ол атадан балаға берілетін категориялар жүйесі арқылы белгілі бір код құрап, дайын күйінде қабылданып отырады. Адамның болмысты тану, түсіну нәтижесінде жинақталған білімі тілдік бірліктерде көрініс табады. Тіл – адамдардың қарым-қатынас құралы ғана емес, әр ұлттың ұлттық болмысын, дүниетанымын, ойлау жүйесін сипаттайтын жалпыадамзаттық құндылық. Ақиқат дүниені тану үдерісінде дүниетанымдық білімдер арқылы жалпыадамзаттық сана қалыптасады. Бұл жалпыадамзаттық сана әлемдегі барлық халыққа ортақ болса, адамзаттың бір-

бірінен ұлт ретінде ерекшеленуі ұлттық сананы қалыптастырады. Демек, тілден жалпыадамзаттық құндылықтармен бірге ұлттық дүниетанымдық ерекшеліктер де көрініс тауып отырады. Әр халық мәдени, тарихи даму барысында қоршаған ортамен тығыз байланыста әлем туралы білімін жинақтап, сақтайды, вербалды сана, ерекше мәдени код түрінде тіл арқылы ұрпақтан ұрпаққа жеткізіп отырады.

Әлемде болып жатқан құбылыстар барлық ұлт өкілдері үшін бірдей қабилданады. Бірақ сол ұлттың ұлттық ерекшелігіне байланысты тілде әртүрлі көрініс табады. Адам тілінде сол ұлттың кәсібіне, мәдениетіне, салт-дәстүріне қатысты тілдік бірліктер көрінеді. Адам – дүние, әлемді, болмысты танушы белсенді субъект ретінде сол дүниенің тілдік бейнесін жасаушы басты тұлға. Тіл – жалпы алғанда, адамның өз болмысы. Сондықтан қазіргі тіл білімінің дамуын гуманитарлық, адам факторына бағытталған, оның өмір сүру ортасына, онда қалыптасқан дәстүр мен нормаға сай өмір сүру әрекетін ескеретін тіл бірліктерінің қызметі мен орнына сол тұрғыдан баға беріп, зерттейтін ғылым ретінде сипаттауға болады.

Адам қандай да бір нәрсе туралы ойлаған кезде оның санасында сол заттың бейнесі көрініс табады. Таңбаның бейнесі ойға негіз болады. Ой сөз арқылы көрініс табады. Сана тіл арқылы көрініс табумен қатар тіл арқылы қалыптасады. Адам санасының қалыптасуы тіл мен сөйлесімнің пайда болуымен тығыз байланысты болады. Адамдар басқалармен қарым-қатынас жасау негізінде тіл мен өзіндік сана қалыптасады. Адам санасындағы болмыстың өзіндік белгілері мен ерекшеліктері білім түрінде жинақталады.

Адам бейнесі тіл семантикасында көрініс табады. Бұл орайда адамның өзі емес, адамның тілдегі бейнесі ерекше маңызды болады. Тілді зерттеу арқылы адамзат өзін өзі танып, түйсінеді. Адам алған әсерін өзі арқылы сипаттайды.

Тіл – адамның дүниетанымын көрсететін ерекше құбылыс, ол этномәдени белгілер арқылы айқындалып отырады. Тілдік категориялар жүйесінде адамның орны ерекше. Осыған байланысты Г.А. Золотова: «Адам – сөйлеуші ретінде де, өзі сөз етіп отырған әлемнің негізгі кейіпкері ретінде де тілдің орталық мүшесі болып табылады» дейді [198, 68]. Тілдің антропоэзектік ұстанымы лексика-семантикалық жүйеде айқын көрініс табады. Бұл бірліктер «мәдени код» ретінде қызмет етеді және сөйлемдердің білімі, стереотипі болып табылады. Мәдениет – күрделі құндылықтар жүйесі және тәсілдері болып табылады, сол арқылы адам әлемді де, өзін де таниды.

Адам өзіне дейінгі жинақталған білім қорын қабылдау арқылы этномәдени қауымдастықтың өмір тәжірибелерін де қабылдайды. Адамның танымдық әрекеті – ақиқат болмысты тануға, қоршаған ортаны түйсініп, жинақталған білімдер негізінде әлемдегі орнын түсінуіне бағытталады. Тіл – сөйлеушінің дүниетаным ерекшеліктерін бейнелейтін құрал, ерекше ұлттық ділдің көрсеткіші, мәдениеттің негізі. Сол себепті М.Балақаев: «Кісі білімді, тәлім-тәрбиені тіл арқылы алады, мәдениетті, өнерді, ғылымды, техниканы тіл арқылы үйретеді. Тіл – ұлт ерекшеліктерінің бірі, ұлт мәдениетінің бір формасы. Сондықтан онда әрбір халықтың ұлттық сана-сезімінің, ойлау тәсілінің, мінез-құлқының нысаналары сақталады. Халықтың сондай ерекшеліктері,

мәдениеті, әдет-ғұрпы, әдеби мұралары, оның психикалық қалпы тіл арқылы ұрпақтан-ұрпаққа ауысып отырады» деп жазады [199, 14].

Сондықтан бағзы заманнан бері ұлттың көне тарихын арқалап, оның ба-рынша рухани-мәдени өмірінен сыр шертіп, ұрпақтан ұрпаққа, атадан балаға ауысып келе жатқан асыл мұра ана тіліміздегі мәдени-танымдық бағытта жазылған ғылыми-танымдық-этнографиялық, когнитивтік бағытта зерттелген жұмыстардың цифрлық базасын жасаудың, ҚТҰК құрамындағы Лингвомәдени ішкорпус әзірлеудің қажеттілігі туындады.

А.Байтұрсынұлы атындағы Тіл білімі институтының А.Жанабекова, Э.Өтебаева, Г.Исаева бастаған Лингвомәдени ішкорпус әзірлеушілері, алдымен, лексикографиялық база жинақтаудан бастады. Олар төрт түрлі бағытта топтас-тырылды:

1. Этнографиялық еңбектер. Арғынбаев Х. «Қазақтың мал шаруашылы-ғы жайында этнографиялық очерк», «Қазақ халқындағы семья мен неке (та-рихи-этнографиялық шолу)», «Қазақ халқының қолөнері», Ахметжанов Қ.С. «Жараған темір кигендер (батырлардың қару-жарағы, әскери өнері, салт-дәстүрлері)», Әбішев Х. «Аспан сыры», Әжіғали С., Байғабатова Н. «Жетісу қазақтарының салт-дәстүрлері мен әдет-ғұрыптары», Әлімбаев Н. «Қазақтың дәстүрлі тіршілік циклі салт-жоралғыларының теориялық мәселелері жөнінде (этносемасиологиялық аспект)», Әнесов Ғ. «Дәстүрлі жерлеу салты», «Нау-рыз: жаңғырған салт-дәстүрлер», Бабалықұлы Ж., Тұрдыбаев А. «Қырандар», Ғабитханұлы Қ. «Қазақ мифологиясының тілдегі көрінісі», Жанпейісов Е. «Қазақ ескіліктері», Ибраев М. «Сарыарқа қазақтарының салт-дәстүрлері мен әдет-ғұрыптары», Қазақ әдеби тілінің сөздігі, Қазақтың Ата заңдары: құжаттар, деректер және зерттеулер, Қазақтың әдет-ғұрыптары мен салт-дәстүрлері: өт-кендегісі мен бүгіні, Қазақтың дәстүрлі киім-кешегі. Ғылыми каталог (қазақ, орыс, ағылшын тілдерінде), Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. Энциклопедия, Қайдар Ә. «*Қазақтар: ана тілі әлемінде (этнолингвистикалық сөздік)*», Манкеева Ж.А. «Қазақ тілінде-гі этномәдени атаулардың танымдық негіздері», Марғұлан Ә. «Ежелгі жыр, аңыздар», Сыздықова Р. «Сөздер сөйлейді», Тәжімұратов Ә. «Шебердің қолы ортақ», Тоқтабай А. «Қазақ жылқысының тарихы», Уәлиев Н. «Сөз мәдениеті», Шойбеков Р.Н. «Қазақ зергерлік өнерінің лексикасы», Шоқпарұлы Д. «Теріден жасалатын бұйымдарды құрастыру технологиясы» және т.б.

2. Ғылыми еңбектер. Ахметов Ә. «Түркі тілдеріндегі табу мен эвфемизм-дер», Жалмаханов Ш. «Түркі лексикасының семантикалық деривациясы», Жартыбаев А.Е. «Орталық Қазақстан топонимиясының тарихи-лингвистика-лық, этимологиялық және этномәдени негіздері», Жылқыбаева А. «Түркі тілдеріндегі ас-тамақ атауларын салыстыра зерттеу», Оразалиева Э. «Қазақ тіл біліміндегі когнитивтік парадигма», Рысберген Қ.Қ. «Қазақ топонимиясының лингвокогнитивтік және этномәдени негіздері», Сәтенова С.К. «Қазақ тілін-дегі қос тағанды фразеологизмдердің тілдік және поэтикалық табиғаты», Смағұлова К. «Мағыналас фразеологизмдердің ұлттық-мәдени аспектілері», Шойбеков Р.Н. «Қазақ тілінің қолөнер лексикасы», **Жанабекова А.** «Қазақ тіліндегі «уақыт» ұғымының тілдік парадигмасы» және т.б.

3. Көркем шығармалар. М. Әуезов, К. Сегізбаев, Ә. Кекілбаев, Қ. Ысқақов, І. Есенберлин, Д. Исабеков, Ж. Аймауытов, М. Дулатұлы, М. Мағауин, С. Мұқанов, Т. Ахтанов, Т. Әлімқұлов, Ш. Мұртаза, Абай, Ы. Алтынсарин, Ә. Нұрпейісов, Б. Майлин, Ғ. Мүсірепов, Ә. Әлімжанов, Х. Бекхожин, Д. Досжанов, Т. Нұрмағамбетов, А. Сүлейменов, Б. Соқпақбаев, Т. Ізтілеуов, А. Сейдімбек, Н. Оразалин, Т. Әбдік, М. Байғұт, О. Бөкеев және т.б.

4. Фольклор. Ертегілер, аңыздар, мақал-мәтелдер, жұмбақтар, жаңылт-паштар және т.б.

Лингвомәдени ішкорпусты әзірлеу барысында Қазақстан Республикасы Президентінің «2004-2006 жылдардағы «Мәдени мұра» мемлекеттік бағдарламасы туралы Жарлығының (2004 ж., 13 қаңтар, №1277) негізінде Үкіметтің «2004-2006 жылдардағы «Мәдени мұра» мемлекеттік бағдарламасын іске асыру жөніндегі іс-шаралар туралы» Қаулысы (2004 ж., 12 ақпан, №171) бойынша академик С.Қасқабасовтың басшылығымен жарық көрген «Бабалар сөзі» жүзтомдық топтамасында кездесетін *діни сөздер, мәдени онимдер, этнографизмдер, кірме сөздер, көнерген сөздер, варианттар* жинақталды.

«Бабалар сөзі» батырлар жыры, діни дастандар, хикаялық дастандар, ғашықтық дастандар, тарихи жырлар, қытайдағы қазақ фольклоры, шежірелік дастандар, көне эпостар, қара өлеңдер, мақал-мәтелдер, балалар фольклоры, ертегілер, қазақ мифтері, Моңғолиядағы қазақ фольклоры, топонимдік аңыздар, күй аңыздар, тарихи аңыздар, өтірік өлеңдер мен мысалдар, түс жору және ырымдар, ауызекі әңгімелер, ғұрыптық фольклор, магиялық фольклор, күлдіргі әңгімелер, ән-өлеңдер, аңыздық жырларды қамтитындықтан, әдеби жәдігерлерді толық түпнұсқа негізінде қарастыруға мүмкіндік туды. Нәтижесінде 100 томнан

- 3640 сөздің варианттары,
- 7551 діни сөздер,
- 8220 кірме сөздер,
- 6020 этнографизмдер,
- 4650 мәдени онимдер жинақталды.

«Бабалар сөзі» жүзтомдығының әр томы Лингвомәдени ішкорпуста *мәтін аты, дереккөзі, аудитория жасы, субкорпус, стиль, мәтін типі, сөзформа саны, мәтін формасы, мәтін көзі, жарық көрген уақыты, белгілеушісі туралы* метабелгі қойылып ҚТҰК базасына енді.

Мәтін аты	Мұхаммед-Қанапия
Дереккөзі	Бабалар сөзі: Жүз томдық. - Астана: "Фолиант", 2004. Т. 10: Діни дастандар. - 2004. - 360 бет.
Аудитория жасы	бейтарап
Субкорпус	поэтикалық мәтіндер
Стиль	көркем әдебиет
Мәтін типі	дастан
Сөзформа саны	2950
Мәтін формасы	жазбаша
Мәтін көзі	«Фолиант» баспасы
Жарық көрген уақыты	2004 жыл
Белгілеуші	Ақтоты Қожахметова

Сурет 25. Метабелгіленім үлгісінің интерфейсі

Ендігі міндет – жинақталған сөздердің мәдени семантикасын ашып, Лингвомәдени ішкорпус материалдарын толықтыру.

Лингвомәдени ішкорпуста қолданушы лингвомәдени бірліктер мен этнографизмдерді белгілі бір тақырыптық топ аясында (адам атаулары, туыстық атаулар, ұлттық тағамдар, ұлттық ойындар, ұлттық киімдер, ұлттық қолөнер, зергерлік атаулар, ыдыс-аяқ атаулары, ұлттық баспана атаулары, қару-жарак атаулары, ұлттық музыкалық аспаптар, ұлттық төсеніш атаулары, киелі сандар) іздей алады және тіркесімділік-семантикалық өрісі, мәдени-семантикалық мағынасы туралы ақпарат ала алады.

«Тіл – өмір шындығы. Сол шындықты уақыт қалдырған қатпарлардан аршып ала білсек, шежіредей сыр шертіп тұрады. Өмірде болған нәрсе тілде де із тастамай кетпейді. Сондай іздің сорабын асыл мұрамыздың тілінен кездестіреміз» [200, 13]. Сондықтан Мәдени-репрезентативті мәтіндер ішкорпусын жасаудың маңызы зор:

- ҚТҰК көлемі бұрынғыдан да ұлғая түседі;
- мәдени мазмұнды мысалдар (қазақ тілінің ауыз әдебиеті үлгілерінен алынған, этнографиялық энциклопедиялардан алынған) табуға мүмкіндік береді;
- гуманитарлық бағыттағы ғалымдардың ғылыми-зерттеу жұмыстарына қолжетімді, тиімді, ұтқыр дереккөзі ретінде қызмет етеді;
- мәдени-семантикалық талдаулар жасаудың әдіснамасын қалыптастыруға жәрдемін тигізеді;
- тілдік-концептуалды, лингвоелтанымдық зерттеулер жасауға мүмкіндік береді;
- тілдік мұра арқылы ұлттық ерекшеліктерімізді толық тануға мүмкіндік беретін негізгі тетік.

8.4 Мәдени-семантикалық белгіленім әзірлемесі

Қазіргі таңда заманауи ақпараттық технологиялар күннен-күнге дамып, барлық ғылым салаларына, оның ішінде тіл біліміне тың серпіліс әкелді. Компьютерлердің жетілуімен лингвистикалық зерттеулерде қолдануға болатын мәліметтер саны, сондай-ақ оларды талдау мүмкіндіктері едәуір өсті. Лингвистикалық зерттеулерде ақпараттық технологиялар қолданылатын негізгі бағыттардың бірі – корпустық лингвистика. Корпустық зерттеулер тиісті мәтіндік деректердің үлкен көлемін талдау негізінде тілді үйрену тәсілі болып табылады. Корпустық деректерді пайдалану белгілі бір мәселелер бойынша статистика алуға, салыстырмалы талдау жасауға және тілдік құбылыстарды кең ауқымда зерттеуге мүмкіндік береді.

Корпус – компьютерде жиналған және талданған үлкен мәтіндер жиынтығы. Корпустық деректер ұлттық және халықаралық корпустарда, сонымен қатар белгілі бір лингвистикалық зерттеу шеңберінде жиналған мәтіндерді камтитын арнайы ішкорпустарда ұсынылады.

Бүгінде Ахмет Байтұрсынұлы атындағы Тіл білімі институтында тілтанымдық ІТ ресурс – «Қазақ тілінің ұлттық корпусы» (әрі қарай ҚТҰК) әзірленіп,

көпшілікке ұсынылды. ҚТҰК-да қазақ тілінің 31 миллион сөзқолданыстан тұратын электронды мәтіндер қоры жинақталған (<https://qazcorpus.kz/>). Ұлттық корпус кез келген тілші-оқырманның тілге қатысты алуан түрлі сұранысын өтеуге дайын. Қазіргі таңда интернет жүйесіне қосылып іздеу арқылы жұмыс жасайтын ҚТҰК бірнеше ішкорпустармен толықты. Олар:

- **Ахмет Байтұрсынұлы ішкорпусы** – А.Байтұрсынұлының ғылыми мұрасы жинақталған әрі кейбір ұғымдардың мәдени-семантикалық мағынасы берілген корпустық база.

- **Ауызша ішкорпус** – электрондық форматтағы ауызша сөздердің транскрибациясы жасалып, сөзжазбасы, сөйлеуші орфоэпиясы мен қалыптасқан орфоэпиялық норма жазбалары берілген корпустық база.

- **Диалект мәтіндер ішкорпусы** – *диалект сөздерді өңір бойынша іздеу жүйесі бар корпустық база. Жергілікті сөздің қай өлкеде, қандай мағынада қолданылатыны туралы ақпарат берілген.*

- **Мақал-мәтел ішкорпусы** – «тірек» сөз арқылы мақал-мәтелдердің тізбегі берілген корпустық база. Мақал-мәтелдердің түсініктемесі (жасалу уәжі, тура және ауыспалы мағынасы) берілген.

- **Тарихи ішкорпус** – тарихи жазба ескерткіштердегі араб және латын графикасында жазылған мәтіндердің корпустық базасы. Мәтіндердің факсимилиясы, тарнскрипциясы, аудармасы берілген.

- **Параллель ішкорпус** – қазақ және орыс тіліндегі көркем әдебиет стиліндегі мәтіндердің теңестірілген аударма мәтіндер базасы. Ішкорпусқа алынған шығарманың екі тілдегі жалпы сөз саны, сөйлем саны, жиілігі сынды статистикалық ақпараттармен толыққан.

- **Ономастикалық ішкорпус** – қазақ тіліндегі топонимдердің (антропоним, ороним, гидроним, ойконим, некроним, фитоним) этимологиясы, семантикасы, географиялық орналасуы жөнінде ақпараттармен қамтылған корпустық база. Топонимдердің сөзжасамдық ерекшеліктері қарастырылған әрі орфографиялық нормасы мен бейнормасы берілген.

- **Жарнама ішкорпусы** – қазақ қоғамындағы бизнес коммуникация негізінде өндірістік қатынастарды реттеуші жарнама мәтіндерінің базасы. Бұл ішкорпус қазіргі қоғамда таратылу орны мен әдісіне қарай *көше, көлік, сауда орындары, БАҚ, интернет* жарнамаларының түрлері бойынша жинақталған.

- **Мәдени-репрезентативті мәтіндер ішкорпусы** – құрамында мәдени-танымдық ақпараты бар тілдік бірліктердің халықтық білімдер жүйесімен анықтауға негізделген корпустық база.

Жалпы алғанда, сөздің лексикалық мағынасымен қатар этномаркерлі мәнділіктерін, этномәдени семантикасын, когнитивтік білімдер жүйесін, прецедентті бірліктерді жүйелеп беру; мәдени-танымдық мағынасы жасырын тұрған тіркестердің астарындағы білімдерді анықтау; көркем шығармалардан жиналған лексикалық қабаттардың мәдени-семантикалық белгіленімін жасау және этномәдени бірліктердің иллюстрациялық (сурет, аудио, видеороликтер, гиперссылкалар) базасын құру, оларды мультимедиялық жасақтамамен қамтамасыз ету мәселелері **ҚТҰК Мәдени-репрезентативті мәтіндер ішкорпусының** басты мақсаты болып табылады.

Мәдени-репрезентативті мәтіндер ішкорпусын әзірлеудің алғашқы кезеңінде зоонимдер, фитонимдер, космонимдер, эргонимдер, абстракттілі сандар, прецедентті есімдер, түр-түс символикасы, ұлттық колөнер, зергерлік бұйымдар, дәстүрлі шаруашылыққа қатысты және Кеңістік пен Уақыт, т.б. байланысты этномаркерлі атаулардың мәдени-семантикалық белгіленімі жасалды. Прецедентті мәтіндер арқылы қалыптасқан тек мәдени мәнге ие атаулардың түрлері мен типтері алынды. Аталған ішкорпуста мәдени тілдік бірліктер қамтылып, олардағы жасырын тұрған кодқа салынған тілдік-мәдени ақпараттар интерпретацияланды. Қазақ тіліндегі ұлттың болмысын, тұрмыс-тіршілігін, мәдениеті мен өзіне тән өмір сүру ерекшелігін танытатын бірліктердің тілдік және мәдени-семантикалық мағынасын ашу барысында материалды талдау, сипаттау және этнолингвистикалық, лексика-семантикалық әдістер қолданылды.

Мәдени-репрезентативті ішкорпустың ғылыми негіздемесіне Қазақ тілінің корпустық лингвистикасының теориялық және практикалық проблемаларын талқылап, алғышарттарын көрсеткен (Жұбанов, 2009; Сүлейменова, 2011; Жұбанов, Жаңабекова, 2017), ҚТҰК-ның алғашқы концепциясын ұсынған (Жұбанов, Фазылжанова, 2014), қазақ мәтіндерін автоматтандыру, олардың метабелгіленімін қою мәселелерін зерттеген (Жаңабекова 2012; Жубанов 2015; Қойбағаров, Мұсабаев, Құлманов, 2012; Жаңабекова, Пірманова, 2019) ғалымдар еңбектері зерделенді [201], [202], [12], [203], [204], [205], [206], [207].

Қазақ тіліндегі этномәдени атаулардың мәдени-семантикалық белгіленімін жасау барысында Әбдуәли Қайдар, Ербол Жанпейісов, Телғожа Жанұзақов, Нұргелді Уәли, Жамал Манкеева, Зәуреш Ахметжанова, Рүстем Шойбеков, Қыздархан Рысберген және т.б. ғалымдардың этнолингвистикалық, лингвомәдениеттанымдық еңбектері арқау болды. Сондай-ақ, лингвомәдени ішкорпустың мәдени-семантикалық белгіленімін [208] беруде профессор Нұргелді Уәлидің идеялық-ғылыми жетекшілігімен жарық көрген 15 томдық «Қазақ әдеби тілінің сөздігінде» үлкен біртөмдік «Қазақ сөздігінде» сөздердің түсіндірмесі мен мәдени анықтамасын беру тәжірибесі басшылыққа алынды. Ұлттық тілдің сөз әлемін жүйелеп берген академик Ә.Қайдардың үш томдық этнолингвистикалық сөздігінің [209], бес томдық «Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі» [210] атты энциклопедияның тоғысатын жері де осы – ұлттық корпус. Мәдени-семантикалық белгіленімнің жалпы сөзтізбесіне енген сөздерге берілген лұғаттық мақаланың құрылымы мен құрылысы, дерексөздерді талғап-таңдаудың жүйесі, олардың мағынасын түсіндіруде аталған еңбектер қолданылды.

Дәстүрлі кітаптан цифрлық сауаттылық деңгейге ауысқан кезде ҚТҰК тілші-ғалымдарға лексика, грамматикамен қоса мәдени-семантикалық мағынаны да жан-жақты зерттеуге мүмкіндік береді. Корпус базасы кез келген оқырманға интернет жүйесінен этномәдени мәліметтердің жинақталған түрін бір жүйеден оңай тауып алуға мүмкіндік беретін үлкен кітапхана деуге болады. Кез келген этномәдени тілдік бірлікке қатысты танымдық әрі қызықты мәліметті қашықтықтан online режімінде оқуға, яғни үйде отырып білім алуға болады. ҚТҰК-ны әзірлеу барысында **мәдени-семантикалық белгіленімді** термин ретінде алғаш рет атап, оның теориялық негізін түсіндіріп берген Анар

Фазылжанова болды. Ғалым мәтіндер базасын қалыптастырудың бастапқы кезеңінде ұлттық мәдени мәні, мазмұны бар мәтіндерге көбірек артықшылық беру керек екенін; этнотаңбаланған тілдік бірліктерде жасырын тұрған мәдени кодтарды ашу арқылы рухани қауіпсіздікті сақтауға және ұлттық мәдениет пен рухтың батыстануға, рухани құндылықты құнсыздандыруға қарсы иммунитет қалыптастыруға болатынын айтты (Фазылжанова, 2016: 105).

Ал енді қазақ халқының білімдер жүйесіндегі этномаркерлі атаулардың мәдени-семантикалық белгіленімін жасау тәжірибесіне тоқталайық.

Мәдени-семантикалық белгіленімге *сөз табы* (зат.е., сан е., етістік.), *стильдік белгі* (этн., діни., көне (бұл сөздер кез келген сөйлеу жағдаятында талғаусыз жұмсала бермейді. Арнайы қойылатын стильдік белгілер арқылы нақты, жеткілікті дәрежеде функционалдық-коммуникативтік ақпарат алады), *мәдени белгі* (заттық және рухани), тақырыптық тобы қойылды және сөз мазмұнындағы ақпараттар түсіндірмеде айқындалып, энциклопедиялық, жалпы сипаттамалық ақпарат берілді әрі халықтық мәдениетпен, руханиятпен, өмір салты, әдебі, әдет-ғұрпымен байланысты жабық кодтағы ақпарат берілді, олардың біршамасы туралы бейнежазба әзірленді.

Белгіленім тәртібі:

1. Сөз таптары: зат есім, сан есім, етістік.

2. Мәдени белгі – заттық және рухани мәдениет.

Заттық мәдениет – мәдениеттегі жеке заттар, құбылыстар мен оқиғалар түріндегі таңбалар жиынтығы. Қолмен ұстап, көзбен көретін заттар ұлттық киім-кешек үлгілері, үй мүліктері, баспана түрлері, төрт түлікке қатысты бұйымдар, азық-түлік өнімдері, тұрмысқа қажетті құралдар т.б. жатады. Яғни Ахмет Байтұрсынұлы айтқандай, «тірнек өнеріне адам ісінен шыққан жасалынды нәрселердің бәрі өнер дүниесі болады, өйткені жасалынды нәрселердің жасалып шығуына адам ақылы, әдіс-амалы, шеберлігі, өнер күші қатысады яғни күн көру ісіне қажет шаруа керек-жарақтары» [211, 150].

Ал **рухани мәдениет** табиғат ісінен шыққан жаратынды нәрселер яғни адамның жан қоштау керегінен шыққан қыз ұзату, келін түсіру, өлік жөнелту, оның артын күту, ит жүгіртіп, құс салу, күнтізбелік, бақсы-балгерлік, мифтік рәсімдермен байланысты қалыптасқан түрлі этномаркерлі тілдік бірліктер, символдық қимыл-қозғалыстар жатады.

3. Стильдік белгі (помета): Лингвомәдени бірліктерді белгілі бір тақырыптық топ аясында іздейді және қолданылуы қарым-қатынас жағдаятына тәуелді болып келетін сөздер арнайы белгілер арқылы көрсетіледі (зергерлік, діни, мал шаруашылық, миф, құқық, тамақ, киім, қолөнер т.б.). Себебі бұл сөздер кез келген сөйлеу жағдаятында талғаусыз жұмсала бермейді. Оқырман арнайы қойылатын стильдік белгілер арқылы нақты, жеткілікті дәрежеде функционалдық-коммуникативтік ақпарат алады.

4. Тілдік (лексикалық) мағына – энциклопедиялық, жалпы сипаттамалы ақпараттар жүйесі. Ашық кодтар мазмұны.

5. Мәдени-семантикалық мағына – халықтық мәдениетпен, руханиятпен, өмір салты, әдебі, әдет-ғұрпы, наным-сеніммен байланысты жабық кодтағы маңызды сипаттамасы. Прецедентті мәтіндер арқылы қалыптасқан тек жабық мәдени кодтан тұратын атауларға берілетін анықтама.

Дегенмен, мәдени семантиканың мағынасын анықтау әдістері корпус жетілдірілген сайын толығып, жаңаланып отыратынан ескертеміз.

Лингвомәдени ішкорпуста қолданушы лингвомәдени бірліктер мен этнографизмдерді белгілі бір тақырыптық топ аясында (адам атаулары, туыстық атаулар, ұлттық тағамдар, ұлттық ойындар, ұлттық киімдер, ұлттық қолөнер, зергерлік атаулар, ыдыс-аяқ атаулары, ұлттық баспана атаулары, қару-жарак атаулары, ұлттық музыкалық аспаптар, ұлттық төсеніш атаулары, киелі сандар) іздей алады және тіркесімділік-семантикалық өрісі, мәдени-семантикалық мағынасы туралы ақпарат ала алады.

Сәукеле сөзінің ұлттық танымдық семантикасы:

СӘУКЕЛЕ баскиім, этн., заттық мәдениет. Дәстүр бойынша қазақ қыздарының ұзатыларда киетін төбесі биік, асыл тастармен әшекейленген сәнді баскиімі. Сәукелені қалыңдық келіншек болғанға дейін, яғни бала сүйгенге дейін киіп жүрген, оның арнайы күмістен жасалған төбелдірігі, маңдайшасы, екі жағында тізілген маржаннан салпыншақ бетжақтауы, алдында бетмоншағы және артқұлағы болады. Жұқа ақ киізден сырылып тігіліп, өнбойы қымбат күрең немесе шымқай қызыл барқытпен қапталып, жиегіне құндыз, кәмшат терілерінен жұрын ұсталады, оқаман оюланып, түрлі-түсті моншақтармен шашақталып, көркем жіптермен кестеленіп, алдыңғы жағы көзді тастармен қапталып, түрлі тағым, алтын, күміс теңгелермен безендіріледі, төбесіне үкі тағылады. Сәукеле мұрагерлікке өтіп отырған. Қалыңдықтың тек өзіне ғана тиесілі меншігі болып саналатындықтан, ол қайтыс болғаннан кейін сәукеле төркініне қайтарылаған. Қайтыс болған келіннің артында ұлы қалса, жасаудан еш нәрсе қайтарылмаған, қызы қалса, жасаудың жартысы қайтарылған. Дүниеден өткен қыздарының белгісі ретінде Сәукелені төркініне қайтару әдеті болған. Өйткені ол арнайы жасалған баскиім, үлкен қуаныш белгісі болғандықтан, төркін жұрты оны марқұм болған қызының көзіндей көріп жүретін ескерткіш ретінде сақтаған. Сонымен, бір ғана *сәукеле* сөзінен көріп отырғанымыздай, қаншалықты терең мән-мағына жатқанын байқауға болады.

Меруертпен шытырмалап, гауһар таспен безеген қызыл мауыты *сәукелесі*, аш белін қынаған алтын оқалы қызыл барқыт камзолының түсін ашқандай жарасып тұр (І.Есенберлин. Алмас қылыш). Басына алтын зерден кесте тігілген биік *сәукеле* киіпті (Ә.Кекілбаев. Үркер). Сөйтсем, әншейінде төбесінен батпандай басып тұратын ханымның ауыр *сәукелесі* көлденең көздің қарап отырғанын көргенде өз-өзінен құстың мамығындай боп жеңілейіп жүре береді екен (Ә.Кекілбаев. Аңыздың ақыры. 2016 ж. 64-б.).

Сәукеле сөзінің тіркесімділік өрісі:

Мақпал сәукеле. баскиім. Мақпал матадан тігілген қалыңдықтың бағалы баскиімі.

Сәукеле байғазысы. кәде. Сәукеле кигізу салтында күйеу жақтан келген құдалардан көрімдікке алатын сыйкәде.

Сәукеле кигізу. Қыздың алғаш рет сәукеле киюіне шашу шашып, байғазы берілетін, кәделермен, бата жасаумен ерекшеленетін салт. Қалыңдық болған қыздың үйіне құда-құдағилар шақырылып, қыз шешесі үкілі тақияның орнына сәукеле кигізеді. Қалыңдық сәукелені ұзату тойынан кейін, күйеуінің ауылына

аттанарда киген. Сәукеле кигізілгеннен кейін қалыңдық болған қыздың енесі және басқа қонақтар көрімдік береді.

Сәукеле көрімдік. Қыз жеңгесі жанына ертіп шыққан қыздардың біреуінің басына сәукеле кигізіп, күйеуден *сәукеле көрімдік* алады. Той басталардан бір күн бұрын қыздың жеңгесі сәукелені киіп, барша сән-салтанатымен, ер-тоқымды жорғамен жүріп маңайдағы ауылдарды ұзату тойына шақырады.

Сәукеле құны. Сәукеле жасауға 5-тен 25-ке дейін жылқы құны жұмсалған. Дәулеті тасыған байлар 100, тіпті 500 жылқының құнына сәукеле жасатқаны туралы деректер кездеседі. Сондықтан да ол жасаудың қымбат бөлігі болып саналған.

Үкілі сәукеле. Төбесі үкінің қауырсындарымен әшекейленген сәукеле. Үкінің қауырсынын тіл-көзден, пәле-жаладан, жамандықтан сақтайды деп жас келіннің сәукелесі мен қыздардың баскиімдеріне қадайтын болған.

ТАБАҚ ыдыс. этн., заттық, Ағаштан, темірден, мыстан, қалайыдан т.б. заттардан жасалатын тамақ ішуге арналған өте шұңғыл емес, жайпақ, дөңгелек ыдыс. Табақтың көлеміне, қолданысына қарай *еттабақ, қонақтабақ, құдатабақ, аттабақ, жайтабақ, тайтабақ, бестабак, төртабақ, үшитабақ, қостабақ, балатабақ* т.б. түрлері бар. Қонақ күту дәстүрінде келген қонақтың жасына, әлеуметтік ерекшелігі мен мәртебесіне қарай тартылатын табаққа сыбағалы мүшелердің қамтылуына қарай *бастабак, ортатабақ, сыйтабақ, аяқтабақ* болып бөлінеді. Ағайын-туыс, жекжат-жұрағатқа, құрметті қонаққа берілетін сыбағалы тамақты *құдатабақ, құдағитабақ, құрдас табақ, жеңге табақ, күйеутабақ, келін табақ, қостабақ, беташар табақ* деп арнайы мүшелер салынған.

Балатабақ. Ас-той, үлкен жиындарда балаларға тартылатын арнайы сыбағалы табақ. Балатабаққа асық жілік, жауырын, ортан жілік, арқа, белдеме, тіл-жақ, қабырға салынады. Баланы бақытқа балайтын қазақ олардың рухани өсіп-жетілуіне де ерекше мән береді. Қонақ болып келген баланы қомсынбай қабылдап, балатабақ тартып ас, дәм сыйлайды.

Бастабак. Үйге сыйлы қонақ, ауыл ақсақалы келгенде, не алыстан өте құрметті, қадірлі қонақ келгенде, не оларды арнайы қонаққа шақырғанда тартылатын сыбағалы, кәделі «төре» табақ ас. Малдың (қойдың) басына жамбас, жауырын, ортан жілігінен және қосымша бауыр, құйрық, белдеме, сүбе қабырға қоса тартылады. Бұған қосымша жал-жая, қазы-қарта және т.б. жеңсік сыбағалар қосылады. Тартылған бастың маңдай құйқасы ұзынынан және көлденеңінен айқас тілініп, танауын төрдегі сыйлы адамға қаратып қояды. Басты кескенде ықылас бата айтылған соң, қонақ үй иесінің ұйғаруымен пышақпен ұртынан, құйқасынан кесіп, өзінен кейінгі адамдарға, үй иесіне ұсынады. Бір немесе екі құлағын бірдей балаларға немесе қонақтың жас серігіне кесіп берген соң, басты үй иесіне қос қолдап ұсынады. Басты жамбас кесуші кеспейді. Бас кесетін адам етке бата берген адам, жамбас кесуші бас дастарқандағы екінші сыйлы адам саналады. Егер қой сойылмай жамбас жілігімен асылса, онда бата істеуші жамбас кеседі. Көзімді қысып қалдым. Үшеуі сасқанынан *бастабак, қорыққанынан қосабак, бір өзіме бестабак* деп тартты (С.Сейфуллин, Таңдамалы шығармалары). – Бізден туған барымыз да, нарымыз да мына

Күмісжан, – деді. Он алтысына дейін ер балаша киіндіріп, асау үйреттіріп, *бастабақтан* қалдырмап едім, одан бұл еркек болған жоқ, құдай бір жаман ұлды маған қиған жоқ (Е.Домбаев, Шығармалары).

Беташар табақ. Беташар атқарылғаннан кейін ет тарту сәтінде келінге бөлек тартылатын табақ. Беташар табаққа үлкендер жағы «ұл тап, тастай бат, судай сің» деп тілек тілеп, ырымдап, ұлтабар мен төстің басын салып береді. Келін бұл табаққа сәлем салып, тағамды қабылдайды. Халық ұғымында көк бауырды «жарыңмен ажырасаың» деп ырымдап, келінге бермеген. Қазақтың келін түсіру салтында, келіннің бетін ашқан жігітке арнап сыйлайтын асты да беташар табақ деп атаған. Беташар табаққа тіл, жақ, жүрек, бауыр, ортан жілік, жауырын, белдеме, арқа, қабырға салып, үстіне таңдай қосады. Табақ үстіне әдемі кестелі қолжаулық жабады, болса күміс жүзік, теңге қояды. Бет ашқан жігіт жолдас-жораларымен отырып, теңгені, жүзікті, қолжаулықты бөлісіп алады да, бата жасайды. Өнші жігіт таңдайым тақылдасын деп, алдымен таңдайды алып жейді. Қалған етті даяшы нөкерлерімен бәрі бірігіп бөліседі. Дәстүрлі ортаның табаққа тіл, жақ, таңдай тартуы «шешен, әнші болындар, бұлбұлдай сайрандар, тілдеріңнен бал тамсын, келіннің беті жарық, жолы ашық болсын», жүрек салғаны «жүрек таза, ниет ақ болсын», бауыр салғаны «бауырдай жақын болындар, сұмпайы тіл болмасын» деген тілекті ғұрыптық сөзбен білдіргені.

Жайтабақ. Ортан жілік пен тоқпан жілік, жауырын, қабырға, омыртқа салынған табақ.

Келінтабақ. Келіндерге арналған сыбаға табақ.

Күйеутабақ. Төс пен асықты жілік салынған күйеу жігіт пен күйеу жолдасқа тартылатын арнаулы табақ.

Қостабақ. Құдалық барысында құдаларға тартылған табақтардан бөлек, артынан арнайы кәде алу үшін тартылған қосымша табақ.

Құдатабақ. Жылқы, сиыр, қой еттерінен мүшелі еттерді топтап жілік, жамбас, омыртқа салынып, құдаларға ұсынылатын сыйтабақ.

Құдағи табақ. Құдағиға арнап тартылатын табақ.

Ошақ табақ. Ошақ басында жүріп, қол ұшын берген, көмектескен кісілерге тартылатын арнайы табақ.

Сыбаға табақ. Сыйлы кісілерге арналған ерекше жасалған табақ.

Табақ тартар. Күйеу жігітке төс салып ұсынылған табаққа сый беру кәдесі.

Табақ тарту. Сойған малдың мүшелерін жіліктеп келген қонақ пен жолы үлкен адамдарға кәделі сыбағасын тарату дәстүрі.

Табақ төңкеру. Ертеде босанатын әйелге көрсетілген емдік шараның бірі. Дәстүрлі халық емшілігінде аққол аналар іштегі бала теріс келген жағдайда толғағы қысқан әйелдің артына табақ лақтырып, «тар құрсағың табақша кеңісін!» деген магиялық мәндегі сөзді айтқан, әрі сол табақты әйелдің құрсағына қойып тұрып онды-солды бұрап теріс келген баланы түзейтін болған. Аққол ананың «оң ба, теріс пе?» дегеніне жанындағы көмекшілері «оң, оң» деп айтқан. Информаторлардың айтуынша, ел арасында кіндік шешеге берілетін сыйды «табағы» деп бір қой көлемінде сый берген (Энциклопедия, 5-том, 2017: 314)

ТОҚАЛ этн. туыстық. Ер адамның бәйбіше үстіне алған екінші немесе одан кейінгі әйелдері.

Ер адамның тоқал алуының бірнеше себептері болған. Біріншіден, ер адам тоқал алу арқылы дәулетін арттыра түсіп, шаруашылығын одан әрі жандандыруды көздесе, екіншісі алғашқы әйелінің, яғни бәйбішесінің шаруашылықты жүргізуге дәрменсіз, ауру, кемтар немесе бала сүймеуі сияқты себептерге байланысты болған. Кейде, бұндай бала көтермеген бәйбіше күйеуіне тоқалдыққа сіңлісін немесе танысын өзінің ұсынысымен алып берген. Үшіншіден, ер адам бірінші әйелін бала күнінде көрмей-білмей әке-шешесі алып бергендіктен оған көңілі толмай, дәулеті арта түскен шағында ел аралап, өзі ұнатқан көрікті қызды айттырып тоқалдыққа алған. Кейде, бұндай көрікті қыздан алған тоқал мен ер адамның жас айырмашылықтары айтарлықтай алшақ болған. Қазақы дәстүрде бәйбішеден кейінгі әйел сәукеле кимеген. Сәукеле – ер адамның бірінші әйелінің мәртебесін көрсететін нышаны (атрибутикасы). Кейінгі әйелдің тоқал атануы – тоқал ешкінің ауыспалы атауы. Әдетте тоқал ешкі деп мүйізсіз ешкіні айтқан. Бір ұл бар тоқалынан Барғана атты, Қоя ма қазақ тұтпай ескі салтты? «Аға өлсе, жеңге мұра» деді-дағы, Қатын ғып Былқылдақпен бірге жатты (Ш.Құдайбердиев, Шығармалары). Ой, тоқалда басың қалсын, тоқал, тоқал деп әлі құрығаныңды білмессің (Ж.Аймауытов, Шығармалары).

Жеңге тоқал. этн. Әмеңгерлік дәстүрі бойынша некелі әйелінің үстіне алған әйелі.

Күң тоқал. этн. Қалыңдықтың қызметшісі болып, оған ере келген басы байлы әйел.

Олжа тоқал. этн. Жаугершілік, шапқыншылық кезінде қолға түсіп, еркектің үлесіне тиген тұтқын қызды некелі әйелінің үстіне алса «олжа қатын» немесе «олжа тоқал» деп айтылады.

Қорыта айтқанда, ҚТҰК-ның **Мәдени-репрезентативті мәтіндер ішкорпусын** әзірлеудің алғашқы кезеңінде қазақ тіліндегі этнографизмдердің мәдени семантикасы, когнитивтік білімдер жүйесі, прецедентті бірліктері жүйеленіп, мәдени-семантикалық белгіленімі қойылды. Ал қазіргі таңда лингвомәдени ішкорпустың екінші кезеңі бойынша, көркем шығармалардан мәдени репрезентативті бірліктерді жинау, оларды лексикалық қабаттар (мәдени онимдер, этнографизмдер, көнерген сөздер, кірме сөздер, діни сөздер және варианттар) бойынша жүйелеп, мәдени-семантикалық белгіленімін жасау және этномәдени бірліктердің иллюстрациялық (сурет, аудио, видеороликтер, гиперссылкалар) базасын құру, оларды мультимедиялық жасақтамамен қамтамасыз ету жұмыстары жүргізіліп жатыр. Лингвомәдени ішкорпустың бұл екінші кезеңі туралы жан-жақты талданған ғылыми ізденістер келесі зерттеуімізде жалғасын таппақ.

Жалпы алғанда, лингвомәдени ішкорпус арқылы қазақ этносының ұзақ этномәдени даму жолын, заттық және рухани құндылықтарын, этнографиялық, мәдени-танымдық ұғымдардың түп-тамырын неғұрлым тереңірек танудың бір құралы ретінде ғана емес, тілдік қазынаның елді тұтастырушылық қызметін сипаттайтын маңызды таным тетігі ретінде құндылығы анықталды. Ұлттың өзін-өзі жеке танып-білуіне, рухани дамуына бастайтын бұл ішкорпустың өнерді,

мәдениетті, кәсіпті, табиғатты тануына және осылар арқылы қоғамымызда өтіп жатқан үдерістерді түсініп-пайымдауға үлкен үлес қосары сөзсіз.

Мәдени-семантикалық белгіленімдер қазақ тілін ана тілінде білетін және екінші тіл ретінде үйренушілерге, қазақстандық жастарға ұлттық коды бар қазақ сөздерінің мәдени-уәждемелік семантикасын игеруге ықпал етеді. Бұл келешекте қазақ халқының ұлттық құндылықтарын, салт-дәстүрін білетін рухани дүниесі бай тілдік тұлғаның қалыптасуына мұрындық болады.

8.5 Мәдени-репрезентативті ішкорпусты эзірлеудегі мәдени бірліктерді аннотациялау проблемалары

Антропоөзекті парадигма аясында сөз байлығының құрылымдық жүйесі ғана емес, танымдық әлемі кешенді зерттеліп, мәдени-репрезентативті тілдік үлгілердің мұрагерлік, қазыналық қызметтері негізінде мәдени-семантикалық мәні ашылады. Сол арқылы кез келген мәдени мазмұнды сөздің бүгінге дейін терең тамырынан желісі үзілмегені айқындалып, соның негізінде әр ұлтқа тән ерекшеліктердің болмысын қалыптастыратын ұлттың мәдени коды танылады.

Осымен байланысты қазіргі тіл біліміндегі антропоөзектік бағытқа негізделген зерттеулерде ғасырлар бойы қалыптасқан адам әрекеті мен күнделікті тәжірибесін тіл арқылы таңбалаған рухани-танымдық ақпаратқа ерекше назар аударылады. Оның маңыздылығы – осындай зерттеулерде айқындалатын мәдени-тілдік репрезенттердің қазіргі қазақ қоғамындағы ұлттық сананың рухани өзегін жаңғыртуға бағытталған заманауи қоғам талаптарына сәйкес дәйектілік қызмет атқаруында.

Тілдің өз бойынан табылатын этностың бүкіл рухани-мәдени байлығының асыл қазынасы іспетті мазмұндық-деректі әлеуеттің жан-жақты ашылып, мәдени репрезентациясының дәйектелуі қазақ тіл білімінде мынадай лексикографиялық еңбектерде жарқын көрініс тапқан: Ә. Қайдар «Қазақтар ана тілі әлемінде» (Этнолингвистикалық сөздік. 1-том – «Адам», 2-том – «Қоғам», 3-том – «Табиғат»; «Қазақ әдеби тілінің сөздігі» (15 том); І. Кеңесбаев. «Қазақ тілінің фразеологиялық сөздігі»; Р. Шойбеков. «Қазақ зергерлік өнерінің сөздігі»; «Қазақ тілінің аймақтық сөздігі»; «Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі (энциклопедия) т.б.

Осы еңбектерге сүйеніп, тіл-мәдениет сабақтастығына орайлас айқындалған, тіл иесінің құндылықтық білімін көрсететін, белгілі бір ұғымдық кеңістікте халықтың мәдени болмыстық ерекшелігін жүйелеп, күнделікті өмірі мен тұрмыс салтын вербалдайтын тілдік құралдың бірі – мәдени-репрезентативті тілдік бірліктер. Қазіргі таңдағы ұлт санасын жаңғырту мен мемлекеттік тіл қызметін жандандыруға қатысты мемлекеттік мүдде бұл істің тіл-қоғам сабақтастығына сәйкес коммуникативтік сапасын (тиімділік, қолжетімділік, ыждағаттылық т.б.) күшейтіп, ұтқырлық технологиясын жетілдіруді қажет етеді. Сондықтан қазіргі тіл біліміндегі антропоөзектік парадигма кеңістігінде құрылымдық жүйе мен жарық көрген лексикографиялық еңбектердің шегі мен мазмұны кеңейіп, заманауи үлгідегі ұлттық корпустық базаны дамытуға бағыт алып отырғаны белгілі. Олай болса, ұлттық корпус базасының сүбелі бөлігін

құрайтын мәдени-репрезентативті тілдік бірліктердің ішкорпусын жасау ісі аса маңызды.

Қазақ тіл білімінде жасақталып жатқан ұлттық корпустың мәденитанымдық тілдік бірліктерінің ішкорпустары құрылымы мәселелерін тиімді жүзеге асыруға ғылыми-әдістемелік негіз болатын В. фон Гумбольдт, А.А. Потебня т.б. ғалымдардың лингвофилософиялық еңбектеріне жалғасқан В.А. Маслова, Е.Телия, Н.Ф. Алефиренко, т.б. ғалымдар мен өз зерттеулерін қазақ тіл білімінде тіл мен мәдениет байланысына ерекше мән берген А.Байтұрсынұлы мен Қ. Жұбанов қағидаларын ұластырған Ә. Қайдар, Р.Сыздық, Е. Жанпейісов, Т. Жанұзақов, Н. Уәли, Ж. Манкеева, З.Ахметжанова, Р. Шойбеков т.б. ғалымдардың тұжырымдары мен талдамалық үрдістері пайдаланылды.

Тіл білімінде ең алғаш рет тіл мен мәдениеттің байланысына қатысты мәселені көтерген В. фон Гумбольдтің: «Среди всех проявлений, посредством которых познается дух и характер народа, только его язык и способен выразить самые своеобразные черты народного духа и характера и проникнуть в их сокровенные тайны» деген тұжырымы, ой-қорытындылары тілді ұлттық феномен деп танитын бүгінгі тіл біліміндегі жаңа бағыттардағы тұжырымдарға негіз болды [212, 69]. Тіл мәдениетпен бір бүтін тұтастық құрып, ажырамас бірлікте өмір сүріп, мәдениеттің тілдің өн бойында қалыптасып, оның дамуындағы алғышарттардың міндетті түрдегі басқышы қызметін атқарады. Осылайша қайсыбір уақытта қандай да бір жерде өмір сүретін этнос мәдениетінің маңызды бөлігін құрайтыны ресейлік ғалым А.А. Потебняның еңбектерінде жалғасса, қазіргі тіл біліміндегі лингвомәдениеттанушы ғалымдар (В.А. Маслова, С.Г. Тер-Минасова, М.М. Копыленко т.б.) шындық дүниені бейнелеуші тілді айнамен салыстырады. Демек, мәдениеттің басқа салаларына қарағанда тілдің өз ортасы бар: «Белгілі бір тілде сөйлейтін кез келген этнос өздерін өзге топтардан сыртқы физиологиялық айырмашылықтарымен ерекшелеп тұратын адамзат қауымының белгілі бір нәсіліне жатады. Тіл мәдениеттен тыс өмір сүрмейді, яғни ол біздің өмір сүру образымыздан хабар беретін тәжірибе-білігімізге, идеяларымызға тәуелді» [212, 85].

Ал Э. Сепир В. Гумбольдтің осындай тіл мен мәдениеттің біртұтастығын ашып көрсететін маңызды әдіснамалық көзқарастарын қолдай отырып: «Мәдениет белгілі бір қоғамның әрекеті мен ойлауының көрінісі, ал тіл – қалай ойлаудың көрінісі. Ендеше, тіл мазмұнды», – деген қорытынды жасайды [213, 190-194]. Осымен байланысты, «Тіл ойды жеткізу мен қарым-қатынас жасаудың құралы ғана емес, мәдени білім аккумуляциясының да басты құралы болып саналады. Күрделі таңбалы жүйені құрай отырып, тіл қандай да бір мәліметті жеткізу, сақтау, қолдану, қайта жаңғырту қызметін атқарады. Мәдениет те тіл секілді белгілі бір хабарды жеткізу қабілеті бар таңбалық жүйе (семиотикалық) болып табылады, алайда ол тілге қарағанда өзін-өзі ұйымдастыруға қабілетсіз, себебі мәдениет ең алдымен – жады, оның басты қасиеті – жинақтау, сақтау және абсолюттік жалпы қауымдастыққа талпыну» [214, 227] деген ой назар аударады.

Осы арада тіл мен мәдениет байланысына қатысты мәдени білімді, мәдени артефактілерді тарату, тіл тұтынушыларға таныту мәселесі өзектеледі. Әрбір

тілді тұтынушы мен қолданушының санасына мәдени білімді енгізу дүниені қабылдайтын ортақ этнотілдік ұжымға тән ұлттық-мәдени кеңістік құруға және тіл жүйесіндегі орталық және шеткері аймақтарға қатысты лингвомәдени коммуникативтік зерттеулерге де байланысты.

Когнитивист-ғалым В.И. Карасик тіл мен мәдениеттің байланысына қатысты «мәnniң тілде кристаллизациясы» туралы айтып, оны «тілдік тұлға санасында дүниенің мәдениетанбалық аккумуляциясы» деп анықтайды [215].

Қазіргі тіл білімінде тілдік бірліктердің мәдени компоненттеріне қатысты Е.М. Верещагин, В.Г. Костомаров, Н.Ф. Алефиренко еңбектері маңызды [216].

Сонымен, жаһандану заманындағы тіл иесінің дүниетанымын байлайтын қазығы – оның бүкіл рухани-элеуметтік, адамтанымдық қазынасы жинақталған тіл деректері. Олай болса, оларды тіл иесінің тарихымен, мәдениетімен, өзіндік ой-санасымен бірлікте қарайтын тілді зерттеудің антропоэзектік бағыты болса, оның саны сапаға ұласқан, кең ауқымды, кешенді, ыждағатты да тиімді тілдік-әдістемелік, танымдық-ақпараттық құралы, заманауи үлгісі – Қазақ тілінің ұлттық корпусы. Оның негізінде мәдениеттанымдық тілдік деректерде астарланған «қазақ әлемінің» бейнесін, ұлттық-мәдени кодын ашу жаһандану заманындағы мәдени сабақтастық, мәдени коннотация, мәдени коммуникация қызметін жүзеге асырып қана қоймай, тілдің коммуникативтік әлеуетін де дәлелдейді. Тілдің терминжасамдық та өрісін кеңейте түседі. Мысалы: тұсаукесер, отбасы, құрылтай, жарғы, нарық, тұғыр т.б.

Осы орайда қазақ ұлттық корпусының мәдениеттанымдық тілдік бірліктерінің ішкорпусына енгізілуі тиіс қазақ тілі қазынасы айшықталған лексикографиялық еңбектерден алынған мәдени семантикалық анықтамалардың үлгілері ұсынылады және 100 томдық «Бабалар сөзінен» алынған кейбір көнерген сөздер, діни сөздер, этнонимдер, этнографизмдер қарастырылды.

Қазақ тілінің аймақтық сөздігі (2005 ж.); Қазақ сөздігі (2013 ж.):

Абылайша (Жамб., Шу). Қос шаңыраққа уықтың өзін ғана шаншып тұрғызатын керегесіз кішкене үй.

Мұны үй туралы этнограф, мәдениеттанушы ғалымдар қазақтың дәстүрлі мәдениетіне, тұрмысына сәйкес, яғни жарықта жүрген кезде, көштің ұзақ жолында дамылдауға қолайлы жеңіл қос түрі деп анықтаған [6].

Бұл атаудың уәжі қазақ тарихымен байланыстылығының дәйектілігі Абылайхан билеген жаугершілік заманда салмағы жеңіл және тігілуінің тез болуына қатысты қалыптасуымен түсіндіріледі.

Адарғы (Қ.Орда, Арал, Сем., Үрж.) – өрмек тоқығанда жіптер шатыспау үшін арасына қойылатын ені бір сүйемдей, ұзыны жарты метрдей тақтайша.

Қазақ қолөнеріне қатысты кілем тоқу үдерісінде өрмек жіптерін екіге ажыратып тұратын құрал. Демек, бұл анықтама мәдени семантикасын ашса, ал оның лингвистикалық интерпретациясы туынды сөздің құрылымындағы ад/аж түбірінің жаңғыртылуымен, яғни тарихи фоно-морфо-семантикалық үдеріс нәтижесімен байланысты [217].

Бұйдагер (Қ.Орда, Қарм.) – көші-қон кезінде көш жайын бақылап, жол сілтеп, көш жолының егжей-тегжейіне дейін біліп отыратын, көш алдында салт атпен жүретін адам. Көшпенді мәдениетке сай күнделікті тұрмыста

қалыптасқан қызмет иесі атауына негіз болған ата кәсіпке сәйкес түйе, өгізді шаруаға қосқанда басқарып отыратын әбзел ретіндегі бас жіп, жетек жіп – бұйда. Осы мағынадан тілімізде оның тізгін – тізгінші, басшы, жетекші т.б. деген ауыспалы қолданысы туындаған.

Ебелек (Қост., Жанг.) – егін машинасының айналып тұратын ағаш қалағы.

Қазіргі қазақ тіліндегі көрсетілген техникалық терминмен қатар «шыр айналған дүние бейнесін көрсететін ою» деген мәденитанымдық атау да бар. Бұл қолданыстардың негізі «ебелек» сөзінің «жеңіл, әлсіз өсімдік» мағынасын білдіретін тура мағынасымен сабақтасады. Халықтың ұжымдық этномәдени санасында қалыптасқан ғалам бейнесінің көрінісіне өсімдіктің бейнелеуіштік белгісі мәдени танымдық тілдік бірліктерге уәж болған.

Жадағай (Шымк., Қызылқ., Сем., Көкп., Ақс., Абай). Астарсыз тігілген жеңіл шапанның атауы ретіндегі бұл аймақтық қолданыстың төркіні де дәстүрлі киім мәдениетінің жүйесіне қатысты. Қазіргі тіліміздегі «жадағай» сөзінің «жайдақ», «жалаң», «қарапайым» сияқты ауыспалы мағыналары да жадағай шапанның атауына жеңіл киім ретіндегі мағынасы уәж болған. Бұл сипат күзгі пальтоның қазақша баламасы ретінде қолдануға негіз бола алатын сияқты.

Салбұрыншы (Шығ. Қаз., Тарб.) – құс салушы, құс ұстаушы.

Қазақтың дегдар өнеріне қатысты аңшы-саятшылардың қыста топталып, бірнеше күнге аң аулауға шығу салтын білдіретін «салбұрын» лингвокультуремасының айшықты суреттелуін М. Әуезовтің «Абай жолы» эпопеясынан көреміз.

Осы сияқты саятшылыққа қатысты «Құсты (бүркітті) қайта қондыру үшін қажетті алдамшы қалта» мағынасындағы «шырға» (Алм., Кег.), «тайдың, құлынның терісінен істелген сырт киім» (Сем., Алм.) т.б. аймақтық қолданыстар – тіл арқылы сақталған мәдениеттің тілдік репрезентациясы.

Ә. Қайдар. «Қазақтар ана тілі әлемінде» (Этнолингвистикалық сөздік. 2013 ж. 1-3-том):

Сөздіктің қазақ тілі әлеміндегі қазақ жұртының рухани мәдениетін танытатын этнолингвистикалық кең ақпараты таңбаланған тілдік қазынасы еңбектің аса маңызды бөлігін құрайды. Соның ішінде адамды о дүниеге шығарып салу, рухына құрмет көрсету өз дүниетанымымыз бен салт-санамызға сәйкес қалыптасқан төл мәдениетіміздің осындай сүбелі еңбектерде дәйектелген көрінісі тілде қалай сақталғаны ұлттық корпустың мәденитанымдық ішкорпусында жүйелі орын табуы тиіс. Солардың бірнеше үлгісін атап өтсек:

Атау кере – ұзақ ауырып, халі нашарлаған адамға арнайы әзірленетін ең соңғы тамағы.

Естірту – өлім туралы хабарлаудың қазақ дәстүрі бойынша өзіндік жөн-жоралғысы.

Қазақтың дәстүрінде суық хабарды ашық айта салмай, психологиялық тұрғыдан санасына жеткізудің өзіндік тілдік қолданысын шебер игерген елдің абыройлы азаматтары, үлкендері бұл істі атқарған.

Көңіл айту – артында қалған жақындарына сөз құдіретін пайдаланып, қолдау, жұбату, үміттендіру қызметіне сай сөздерін айту.

Жоқтау – көбінесе елге белгілі тұлғалар өмірден өткенде сөз арқылы зарлы жыр түрінде айтылатын қастерлі дәстүр.

Бұл дәстүр марқұмның жақын-туыстары мен жұртының жадында сақталатын «сөзбен тұрғызылған ескерткіш» іспетті рухани дүние ретінде танылады.

Сүйекке түсу – қаза болған адамның мәйітін арулау рәсімі.

Марқұмды жөнелту рәсіміне жанына жақын, тірлікте жақсы сыйласқан адамдар кіретіні де қазақ дәстүріндегі тәртіп ретінде қалыптасқан.

Тұлдау – марқұмның тақымы тиіп, өзі мінген атының жал-құйрығын күзеп, ешкімге мінгізбей, жылқыға жіберіп, семіртіп, өзінің асында сою.

Осы қатарда келтірілген «тұлдау» рәсімінің ұлттық корпустағы мәдени-тілдік репрезентациясының мәдени-семантикалық мазмұнын ашатын тілдік ақпарат арқылы қазақтың дәстүрлі мәдениетінде көрініс тапқан жылқының қазақтың ұлттық болмысы мен мінез-құлқын айқындауға септігін тигізеді. Көшпенділердің архаикалық санасын бейнелейтін батырлар жырында, паремнологиясында жылқы тек көлік қызметін атқарушы ғана емес, ол ер қорғаны, жолдасы ретінде көрініс табады. Сондықтан да батырлардың серігі іспетті Байшұбар, Тайбурыл, Ақбозат т.б. жылқылардың аттары антонимдер сияқты прецеденттік атауларға айналған. Сондықтан ат иесі өмірден өткеннен кейін жасалатын «тұлдау» рәсімінің негізінде жылқы «әруақтар әлемі мен адамдар арасында медиаторлық қызмет атқарған деген түсінікке сай мифтік таным қалыптасқан. Жалпы, қазақ халқының қоғамдық-әлеуметтік, рухани-мәдени өміріндегі жылқының қозғалыс, өмір сүру көзі, ел тану, ұланғайыр кеңістікті игеру құралы ретіндегі бейнесі тілінде жан-жақты көрініс тауып, қалыптасқан кешенді мәдени жүйе құрайды. Қараңыз: жылқының жасына қатысты: құлын, жабағы, тай, құнан, дөнен, бесті т.б.; қозғалысына қатысты: аяңдау, желу, жорту, шабу т.б.; жынысына қатысты: байтал, бие, сәуірік, айғыр, азбан т.б.; түсіне қатысты: қыран, баран, ала, шұбар, қасқа, кер, боз т.б. Жылқы ер-тұрманының әрбір бөлшегі жіктеліп, аталған: тебінгі, тоқым, терлік, айыл, ауыздық, тізгін, сағалдырық, құйысқан, жүген, өмілдірік т.б.

Тілде көрініс тапқан жылқының қазақ тұрмысындағы жан-жақты қызметі қазақ халқының этномәдени санасында ерекше мифопоэтикалық жүйе құрған ғаламның кез келген ұлттық, тілдік бейнесінің әмбебаптығын және өзіндік ерекшелігін айқындайтын метафора, теңеу және символ болып табылады [219].

Мысалы, теңбіл көк, ақ бөрте, тарлан боз, құла қасқа, ал қара көк, ақ табан, сұр жорға, қара қасқа, жылтыр қара, сүліктей қара, тобылғы торы, қара кер, құла жирен, боз, шұбар, мақпал, қара, әсем қоңыр, көк шолақ, кербесті, ақ бақай.

Қазақтың дәстүрлі мәдениетіндегі жылқының этнокогнитивтік санадағы болмысы мен мәдени антропологиядағы параллель де осыны көрсетеді («Байтал шауып бәйге алмас» т.б.). Сол сияқты ол халықтық метрологияда да (бие сауым уақыт, бие байлар кез, ат төбеліндей, бес биенің сабасындай т.б.), салт-дәстүрлердің қалыптасуында (бәйге, көкпар, қыз қуу, ашамайға мінгізу т.б.), ырым-ғұрыптарда (ақ боз ат шалу, бозқасқа атау т.б.) көрініс береді.

Демек, қазақтың дәстүрлі мәдениетіндегі жылқының заттық және рухани құндылықтар жүйесіндегі орнын тіл арқылы дәйектеу мәдени-репрезентативті ішкорпустың сенімді дерегі болмақ.

Ұлттық корпусқа этномәдени мазмұнды тілдік бірліктердің мәдениетанымдық репрезентациясына негізделген ішкорпустарын енгізу тіл арқылы ұлт болмысын таныту мен сана жаңғырту мәселесіне қатысты аса өзекті. Оны жүзеге асырудың тиімді үлгісі қазақ білімінде жеткілікті (Ә.Қайдар, Р. Сыздық, Е. Жанпейісов, Н. Уәли т.б.). Бұл мәселенің кешенді сипаты оның саялық білім қоры ретінде тарихи-мәдени, этнографиялық, археологиялық т.б. зерттеулердің ақпараттық деректерін тиімді пайдалану қажеттілігін көрсетеді.

Мысалы, ұлтымыздың мәдени болмысын, мұрагерлік, коммуникативтік, прагматикалық тұрғыдан танытатын коммуникативтік базасының заманауи да кешенді үлгісі ретіндегі ұлттық корпусты дамытуда, ішкорпустарды жетілдіруде дәйектік көздердің тілдік деректерін кешенді зерттеу үдерісіне ерекше көңіл бөлінбек.

Сондай көздердің бірі 2013 жылы М. Әуезов атындағы әдебиет және өнер институты жариялаған «Бабалар сөзі». Осы мақалада «Бабалар сөзінің» тілдік дәйегі ұлттық корпусқа деректік көз ретінде алғаш рет ұсынылып отыр [220].

«Бабалар сөзінің» «Ғашықтық дастандар» бөлігінен алынған мәдени-репрезентативті тілдік бірліктер де ұлттық корпустың ішкорпусын дәйектеп, біздің ойымызша тіл арқылы ұлт болмысын, ғасырлар бойы қалыптасқан төл мәдениетін дәлелдейтін рухани көздердің бірі ретінде болашақта жан-жақты қарастырылуы қажет.

Этномәдени мазмұнды тіл бірліктері:

Наурыз. Қазақтардың бағзы мәдениетінің даму үдерісінде соғды, иран, хорезм мәдениетінің ықпалы, әсері болғаны белгілі. Қазақтың күнтізбесінде, ай атауларында ол көрініс тапқан. Соның бірі – Наурыз айының атауы. Иран тілінде «ноу» – жаңа және «руз» – күн дегенді білдіріп, әрі қарай барша түркі тілдерінде «жаңа жылдың алғашқы күні», «жыл басы», «Ұлыстың ұлы күні» т.с.с. мәдениетанымдық ұғымдарды білдіретін сөз тіркестерінің қалыптасқаны белгілі.

«Ғашықтар дастандарының» дерегіне сүйенсек: Бір күні наурыз берді жұрттың бәрі, шат болып жиылыста жас пен кәрі.

Қамшы – қазақ халқының материалдық мәдениетіне қатысты байырғы кездегі шаруашылық, мәдени өміріндегі маңызды ер-тұрманның бірі. Сонымен бірге ол қолданыс барысында этномәдени жүйенің бірлігіне де айналып, әлеуметтік-құқықтық, этикалық, мифтік т.б. мәнге де ие болғанын тілде қалыптасқан сөздер мен сөз тіркестері айғақтайды: қамшы тастау, қамшының сабы сынды, қамшы таянды т.б.

Осының бәрін кешенді талдап, қамшы сөзінің мәдени семантикасын ашып, концептілік жүйеде қарастырған ғалым Н. Уәли қамшы сөзінің лингвистикалық реконструкциясын да анықтайды: қам – бақсы, шы – сөз тудырушы қосымша [221].

«Ғашықтар дастанында»: Қамшымен ашуланып салғанында, Бір асқан батырыңнан қолы қатты.

Діни сөздер. Бұлардың дастанда кең қолданыста және бірнеше дыбыстық нұсқада айтылатыны байқалады: Құдай, Алла, пайғамбар, Ислам, неке, Құран, намаз, хадис, Қағба, садақа, сүннет, тамақ және Алла-Аллаһ-Алда; Құдай-Құда-Құдай-ай; молда-молла т.б.

Этнонимдер. Ғашықтық жырларда қолданылған көптеген пайғамбар, адам аттары т.б. мәдени танымдық даму үдерісінде прецеденттік этнонимдерге айналып кеткені белгілі: Мұхаммед, Сүлеймен, Ұлықпан, Ескендір, Жүсіп-Зылиқа; елді-мекен, жер-су атаулары: Йемен, Қытай, Һүндістан, Мысыр, Гүлстан; Сахара, Алтай, Хорасан, Қобда, Иранбақ, Ніл. Сол сияқты сол кезеңнің мәдени-рухани орталықтары туралы ақпараты таңбаланған қала аттары: Құдыс, Үрім, Бағдат, Басра, Түркістан т.б.; этнос атаулары: қытай, орыс, араб, керей, зәңгі т.б.; мифтік атаулар: дию, пері, періште, мыстан кемпір.

Көне сөздер. Әйбатлап (Әйбатлап бір көшкіні жасадылар, Көйлекні бір темірге қашадылар). Қазіргі қазақ тілінде әйбат сөзі аймақтық қолданыста.

Сол сияқты: түмен, көшік, үшбу, кеп, тәрік, сейір, көнек, пұл, ділла, канадил, мейман т.б. мәдени атауларды көнерген сөздер қатарында қарастыруға болады.

Ал, осындай ежелгі қисса-дастандарда кездесетін хауыз, әуіз, әуез, зияпат, кебаб, пұл сынды көнерген сөздердің қазіргі тілімізде жаңғырып, қазіргі күнде белсенді қолданысқа түсуін қазіргі қоғамдағы рухани жаңғыру үдерісіне сай мәдени танымдық трансформация, коммуникация нәтижесіндегі сәтті, жаңа қолданыстардың, терминдердің қазақша баламасының кумулятивтік үлгісі деп санаймыз. Соның нәтижесінде қазір олар хауыз (бассейн), әуез (мелодия), зияпат (банкет), кәуап (шашлык) т.б. мәдени танымдымызды кеңейтіп, жалғастырып, тілдің мәдени танымдық әлеуетін жүзеге асыруда.

Болашақта 100-томдық Шығыс ауыз әдебиеті үлгілерінде кездесетін мифтік танымға сәйкес мәдениеттің архаикалық үлгілерінің тамырын терең кешенді мәдени-рухани көз екенін айғақтайтын мифолог-ғалым С.Қондыбайдың шығармаларының мәдени-тілдік деректері осы мақсаттағы кешенді деректік дәйек ретінде қарастырылуы тиіс.

Тіл – өз бойында мәдениеттің арқауын құрап, негізін қалайтын, ұлттық рухты айқындап тұратын ұлы күш, мәдениетімізді танытушы бірден-бір қуатты құрал, ұлттық мәдениет тірегі. Төл мәдениетіміз сақталған негізгі тарихи-рухани көздердің бірі.

Тіл арқылы сақталған, қалыптасқан этномәдени атаулар – ұлттық танымдымызды айғақтап, ұрпақтан ұрпаққа жеткізіп отыратын рухани байлығымыз. Соны бейнелейтін мәдени-репрезентативті тілдік бірліктер ұлттың өткені жайлы мол ақпарат береді. Тарихи-рухани құндылыққа ие мәтіндерді оқу барысында сол кездегі тіл иелері өмір сүрген кеңістіктегі саяси-әлеуметтік жағдайды, болған түрлі тарихи, мәдени, әскери, экономикалық оқиғаларды, соған қатысқан кейіпкерлерді, тұлғаларды, оқиға болған өңірдің түрлі мәдени ерекшеліктерін тіл арқылы тануға болады. Оның себебі – тіл иесінің болмысты танып, түсінуі өзі өмір сүретін ортада жинақтаған білімі мен тәжірибесіне сәйкес түзілуінде. Сондықтан заманауи лингвистикада тілді қарым-қатынас жасаудың тәсілі деп қана емес, жалпы мәдени білімнің жиынтығы деп қарайтын қағида қалыптасқан. Осымен байланысты академик Ә. Қайдардың тілді зерттеушіні «тіл тамыршысы» деп санауы тіл-мәдениет сабақтастығын айқындаудың кешенді болмысын көрсетеді.

Себебі ұлттық мәдениет көрсеткіші, тілде сақталған және қалыптасқан ұлттық мәдени код ретіндегі мұндай мәдени-репрезентативті тілдік бірліктер

мемлекеттік тілдің жан-жақты қоғамдық қызметінің дамуына, баюына, оның тұтынушыларының сұранысына ие болу мен ел мүддесіне лайық зор үлес қосары сөзсіз. Бүгінде ғасырлар қойнауынан шығарылып, күнделікті тілдік қолданысқа енгізіліп, жалғасып жатқан мұндай сөздердің қолданыс аясының кеңейіп, көбіне терминдік мәнге ие болып жатқаны – соның дәлелі.

Қазіргі қоғамдағы ұлттық санамызды жаңғыртып, халықтық білім деңгейінде қалыптасқан мәденитанымдық санамызды дамытатын сенімді көзі ретінде заманауи рухани-коммуникативті құрал ұлттық корпус танылып отырса, оны кеңейтіп, жетілдіріп, дамытушы – сайып келгенде, ұлттық корпус базасының өзекті ішкорпусын дәйектеп, жетілдіретін мәдени-репрезентативті тілдік бірліктер жүйесі.

8.6 Қазақ лексикасындағы жаңғырған атаулардың мәдени-семантикалық сипаты

Тәуелсіздік алғаннан бері күні бүгінге дейін қоғамда болып жатқан түбегейлі өзгерістердің нәтижесінде бұрынғы уақытта жұрт жаңғыртты (өріс, қоныс ауыстырды), сүйек жаңғыртты (қайтадан жекжат болды), төсек жаңғыртты (қандай да бір себеппен екінші рет үйленген адам туралы) тәрізді сөз тіркестерінің «жаңғыру» ұғымымен байланысты қолданыстың аясы кеңейіп, төмендегідей жаңа қолданыстар пайда болды: рухани жаңғыру, тарихи-мәдени ескерткіштерді жаңғырту, жаңа тұрпатты жағырту, тарихи сананы жаңғырту, толыққанды жаңғырту, жаңғырту жолы, ұлттық жаңғыру, жаңғыру ұғымы, аумақтық жаңғыру, заманға сай жаңғыру, түбегейлі жаңғыру, жаңа жағдайда жаңғыру, музыкалық дәстүрлі жаңғырту, саяси жаңғыру, жаңғырған ел, жаңғырту ісі, жаңғырған қоғам, жаңғырған атаулар, жаңғыру үлгілері, экономикалық жаңғыру, жаңғырудың алғышарттары, қайта жаңғыру, жаңғыртылмалы несие және т.б. Бұлар қазіргі ақпараттық технологиялар тасқыны мен жаһандану дәуіріндегі тарихи-элеуметтік өзгерістерді жан-жақты бейнелейтін «өзгеріс», «жандану», «қайта қалпына келтіру», «қайта тірілту», «жаңаша сипат беру», «толығу», «түлеу», «үйлесімді дамыту», «жаңару», «жақсару», «кұлпырту», «қайта қолдану» сияқты сан алуан мағыналарды білдіреді. Осыған байланысты тіл білімінде жаңа атаулар, жаңа қолданыстар, жаңғырған атаулар, қайта жаңғырған атаулар тәрізді сөз-терминдер қолданылып келеді. Бұлардың ұғымы, мағынасы жақын болғанымен, өзіндік айырмашылықтары да бар.

Академик Р.Сыздық «тілдің бүгінгі лексикалық қазынасындағы көне, ескі, жаңа деп танытуға болатындарын табу, талдау, таныту» маңызды деп көрсетті. Бұл тұжырым айрықша мәнге ие. Өйткені қазақ әдеби лексикасының қазіргі дамуы мен жетілуі сипаттарында байқалатын жайттың бірі – көне сөз деп аталып келген бір лек тілдік бірліктердің қоғам дамуындағы өзгерістерге сәйкес прагматикасының артуы, актуалдануы. Айталық, ауызашар, неке қию, саудагер, намазхана, сүндеттой, шапан жабу тәріздес мағынасы жағынан «бірі таныс, бірі бейтаныс» діни, этнографиялық мазмұндағы тілдік бірліктердің, мәдениет пен қолөнерге қатысты атаулардың қолданысы артты, олар ұлттық болмысты тіл арқылы таныта түсті. Осындай тілдік бірліктерді жаңғырған атаулар деп

бағалауға болады және олардың жаңа сөздер ұғымына қатысы бар. Қазіргі қазақ тілінде жаңғырған атаулар бүгінгі ақпараттық технологиялар тасқыны мен жаһандану дәуіріндегі тарихи-әлеуметтік өзгерістерді жан-жақты бейнелейтін «өзгеріс», «жандану», «қайта қолдану», «қайта қалпына келтіру», «қайта тірілту», «жаңаша сипат беру», «толығы», «түлеу», «үйлесімді дамыту», «жаңару», «жақсару», «құлпырту» сияқты сан алуан мағыналарды білдіреді.

Жоғарыдағыларды жинақтай айтқанда жаңғырған атаулар дегеніміз сөздік қорымызда болғанмен бір кездерде сирек айтылатын болып, қолданыстан шыға бастаған, мағынасы күнгірттенген яғни ұмытылған, бірақ қазіргі кезде ұлтымыздың жаңғыруына байланысты қайта қолданылып жүрген жаңа мән-мағына ала бастаған сөздер мен сөз тіркестері. Оларды кейде жаңа сөздер деп қолдану шартты, өйткені бұрын естімеген немесе бір жерден оқымаған адамның ондай сөздерді бұрын болмаған, тек соңғы кезеңде пайда болған сөз екен деп қабылдауымен байланысты.

Жаңғырған атаулардың тағы бір тобы 2006-2013 жылдар аралығында он бес томдық «Қазақ әдеби тілінің сөздігінде» қамтылған. Алайда, қандай түсіндірме сөздік болса да, онда тілдегі барлық сөзді түгел қамту мүмкін бола бермейді. Он бес томдық түсіндірме сөздіктен ширек ғасырдай бұрын жасалған 10 томдық «Қазақ тілінің түсіндірме сөздігіне» байланысты зерттеуші әрі осы сөздік авторларының бірі Б.Қайымованың: «Осы уақытқа дейін тілдің сөз байлығын толық қамтыған бірде-бір сөздік ешбір елде болмағаны тарихтан белгілі. Арнайы зерттеген ғалымдар оған нақты дәлелдер арқылы көз жеткізіп отыр. Сондықтан қазақ тілі топырағында алғаш тәжірбие саналатын бұл сөздік ондай кемшіліктен құр емес. Қазір байқап отырғанымыздай, көптеген сөздер сөздікке енбей қалып қойған...» деген пікірі жоғарыдағы пікірімізді нақтылай түседі (Қайымова, 1983: 7).

Өзіміз өмір сүріп отырған әлемді адам-қоғам-табиғат деп топтастырсақ, бұлардың әрқайсысы да жаңғырып жатқан атауларда көрініс тауып отыр. Мәселен, «табиғат» деген ұғымға кіретін өсімдіктерді алып қаралық. Қазақ тіл білімінде өсімдік атауларын зерттеуші Б.Қалиевтің және т.б. зерттеушілердің еңбектерінде мольнан қамтылған. Аталып отырған зерттеушінің 40 жылдан астам уақытын жұмсап жинаған материалы жеке сөздік болып та шықты. Солай болғанмен, тәуелсіздік жылдарында баспасөз беттерінде жарияланып келе жатқан өсімдік атаулары аз емес... Бұны тіліміздің байлығының көрінісі деп бағалауымыз керек.

Халық емшісі Сейітқамза Қалиев «Ана тілі» газетінің бетіне жарияланған «Өз тіліміздегі өсімдік атауларын білесіз бе?» деген мақаласында былай деп жазады: «Он томдық қазақ тілінің түсіндірме сөздігінде » бар болғаны 768 ғана өсімдік атауы бар. Соның өзі кейбіреуі жаңсақ жазылып, атаудың толық мағынасын бере алмайды. Демек, қазақ жерінің көптеген өсімдік түрлері басқа ұлттардың, әсіресе, орыстардың атауымен тұрғаны намысымызға сын деп ойлаймын. Аталған түсіндірме сөздікте 67 мың сөз тіркелген екен. Оған басқа жақтан іздемей-ақ, өзімізде өсетін өсімдіктердің атауын қоссақ, сөз қорымыз қанатын жайып, қомдана түсер еді. «Ертеректе Тарбағатай өңірінде өмір сүрген Оспан Бабалықов деген ақсақал өсімдіктің он мыңнан астам түрінің атын атап,

түрін түстеп беруші еді» деп еске алып отыратын күні кеше ғана көз жұмған ғұлама қария Жағда Бабалықов.

Осы орайда еріксіз ойға қаласыз. Осыншама өсімдік болмаса, соншама атауды сол адам қайдан білді. Оны кезінде ескеріп, қағазға түсіре алмағанымыз өкінішті-ақ. Ағылшындардың сөздік қорында 415 мың, орыстарда 120 мыңнан астам сөз болса, осының аса қомақты бөлігі өсімдіктердің атауымен көркейіп тұрған жоқ па?

Әлі күнге дейін ел ішінде ауызша айтылып, тұрмыста қолданылып жүрген өсімдік атаулары түсіндірме сөз қорымызда жоқ болуы немесе басқа мағынамен айтылуына бірер мысал келтірейін. Түсіндірме сөздікте «сасық шөп» деп аталатын шөпті ел ішінде «гүлбуын» деп те атайды. Оны сасық дегеннен гөрі гүлбуын деген мағынасы құлаққа жағымды әрі айтса айтқандай әр жапырағының түбінен гүл қауашақтары ашылады. Ол – жүрек дәрісі ретінде атына заты сай көрікті өсімдіктердің бірі, – дей келіп, қазантолды, күйдіргі және т.б. өсімдіктердің қасиеттері жайында сыр шертеді. Автордың айтуынша, ұзақ жасайтын бүршікті сібір қарағайының дәрілік қасиеті бар, уақыты келгенде өзі үзіліп түсетін, дәнінің піскен мезгілі ел ішінде қазантолды деп аталса, күйдіргі деп айтылатын дәрілік өсімдік буын ауруларын емдеуге қолданылған. Оның емдік қасиеті туралы емшінің өзі: «Оны денеңізге мерзімінен артық ұстасаңыз күлдіретіп күйдіріп тастайды. Ал шамалы ғана ұстасаңыз, буын ауруларын жазады. Екі тізесін баса алмай қалған адамдардың осы емнен соң толыққанды жазылып кеткенін көзбен көрдік», – дейді (Қалиев, 2010: 5).

Жаңғырған атауларға байланысты «Қазақ лексикасындағы жаңғырған қолданыстар» атты кітапқа жазған алғысөзінде Р.Сыздық былай деп жазады: «Жаңа қолданыстар тек жаңа ұғымдарды білдіру үшін ғана емес, бұрыннан келе жатқан зат, бұйым, құбылыстарды дәлірек, дұрысырақ етіп қайта атау үшін де туып жатады... Қысқасы, бүгінгі таңдағы оқырманға күнделікті баспасөз беттерінен, радио, телехабарлардан, ғылыми әдебиеттен, тіпті көркем әдебиет кітаптарынан жүздеген соны сөздер, тосын тіркестер немесе байырғы сөздердің жаңа мағынада жұмсалып отырғаны жақсы сезіледі» дей келіп, олардың пайда болуы мен атқаратын қызметі туралы: «Бұл жаңа элементтер қайдан, қалай жасалады? Бұл жердегі «жаңа» деген анықтауыш-атрибут шартты. Әдетте жаңа сөз дегеніміз – тілде бар сөздердің (түбірлердің) жұрнақ, жалғау, бір-бірімен қосақталу (бірігу, кірігу, жұптасу), қысқарып барып бірігу (аббревиатура) мағына ауыстыру сияқты амалдардың бірі арқылы құбылған түрлері. Біз теріп, тіркеп отырған «жаңа сөздер» де бұрынғының «жаңғырықтары» емес. Бірақ олар естіліп барып, өшіп кететін дыбыс жаңғырықтары емес, материалдық дүниелер – тіл единицалары, яғни белгілі бір ұғымды білдіретін дербес сөздер (жаңа тіркестер), сөздік қазынаны молайтатын, тілдің ілгері даму процесін көрсететін бағалы дүниеліктер» деп жазады (1985: 3-7).

«Қазақ әдеби тілінің сөздігінде» (5-том, 340 б.) ерен деген бұта атауына «денесінен от алғыш эфир майын бөліп шығаратын бұта» деген түсінік беріледі. Ал Қозыбай Құрманның аталуы осыған жақын ерен деп аталатын ағаш туралы «Шарын шатқалы» деген мақаласында ерenniң сирек кездесетін, тек Американың Невада штатындағы «Гранды» шатқалында және осы біздің

Шарында (Алматыдан 200 км жерде) өсетін темірдей қатты ағаш екенін айтады.

Автордың өз сөзімен айтқанда: «Тоғай демекші мұнда тамырын тарихтың қатпар-қатпарына сүңгітіп-сүңгітіп алып, жаһандық мұз дәуіріндегі сұрапылда да сұламай өткен ежелгі соғдылық ерен ағашы бар. Сұламақ түгілі шатқалдың түкпірінде әлі шарапаттана жайқалып тұр. Арасында діңіне жеті адамның қол ұстасқан құшағы әрең жететін алыптары да ұшырасады. Осындайда бір айта кетерлік жайт, бұл ағаш ел аузында «шаған», «шетен», «шеген», тіпті «шынар» деп те аталып кеткен. Ал бұл туралы көрнекті журналист Жанболат Аупбаев: «Онда (Шарын шатқалында) темір сияқты өте қатты, ыстық температураға шыдамды ағаш өседі. Оны биолог-ғалымдар Шарындағы шетен (яшень) тоғайы деп жүр. Біздіңше, бұл дұрыс емес, қате атау. Себебі шетен ағаштардың ішіндегі ең жасығы. Оның әлсіздігі сондай, егер қазық жасап жерге қақсаңыз, жартысынан морт кететін, ол соққы тиген сайын бой-бой болып сөгіліп, шытынап кетеді. Бала күнімізде Сарытоғайдан әкелген жоғарыдағы темір ағашты аталарымыз бен әкелеріміз: «Бұл – ерен ағашы. Оны өзен көпірінің астындағы тіреуішке қойсақ шірімейді, үйдің төбесін жабарда белағашқа пайдалансаң сынбайды» деп отыратын» деп айтып, былайғы жұртқа бірнеше рет ескертпе жасаған» (Құрман, 2016: 12).

Қазіргі қоғамдық-әлеуметтік салада, шаруашылықта, салт-дәстүрімізде, діни ғұрыптарда, қолөнерінде қайта қолданылып жүрген немесе көпшілікке таныс емес атаулар рухани дүниемізді жаңғыртатын тілдік қазынамыз. Олардың бәрін бірдей көпшілік біле бермейді. Мысалы: тобар – көктемде мал төлегенде таракпен тарап алатын, одан балалар доп жас жасайтын сиырдың жүні. Тегелдірік – атқа мінген адамның етігі, балағына аттың тері жұғып былғанбасын деп тоқымға бекітілетін былғары ер жабдығы (Қыдырбекұлы, 1991: 3). Айілгек – әйелдердің камзол, шапан тәрізді үй киімдері мен сырт киімдеріне тағатын мүйізден әшекейленіп жасалған, ай орағына ұқсас, көлемі баланың саусағындай сүйектен жасалған түйме түрі. Арқан тарту – этн. Арқан жасауға арналған тіндерді ширатып, арқан дайындау тәсілі. Сұламажелі қолөнер. Ағын суды бейнелейтін ою-өрнек сабағының қайталанып отыруы. Баспатігіс қолөнер. Тігістің бір түрі. Бұл тігіс көбінесе өңіржиектің астына астар салып баспалап тіккенде қолданылады. Сонымен қатар әжелер жаулықтарының немесе кимешектерінің жағын, өңірін қиғаштап кестелегенде баспатігіс әдісімен тігеді. Құрымдау қолөнер. Теріні арнайы ыс күйе салып қайнатып дайындарын суға батырып қара түске бояу. Шапжіп этн. Түйеге артылған жүктің орнықты тұруы үшін шаптан тартып байланатын жіп. Ақтық қою этн. Қазақтың дәстүрлі өлік жөнелту, аза тұту салтында орын алатын рәсім – мата түрінде үлестірілетін рәсім. Өрнекжелі қолөнер. Жеке бір өрнек түрінің айқасып немесе кезектесе түсуінен жасалған жасалған желі. *Шырлық* – егістіктен алынған өнімнің күнкөріс үшін қалдырылатын ең қажетті мөлшері (минимум) (Энциклопедия; 2017: 723). Жүкбұзар / жүкшешер этн. Жасаудың жүгін шешкен адамға ұзатылған қыз төркіні жағынан берілетін кәделі сый. Дәужыр – баптауды онша көп қажет етпейтін күрішке арналған танап, жер. Күмбей – күріштің әбден пісіп жетілген кезі. Талдырғы – арықтың бойымен қатты келе жатқан суды

шөп-шаламмен бөгеу (Бектұров, 1984: 13). Садақа – «тек ақша беру, мал шашу емес, қайырымды іс қылу. Айталық, бір қария көшенің келесі бетіне өте алмай жатса немесе іздеген көшесінің атын ұмытып қалған болса, тағы да басқа жағдайларда қол ұшын беру, демеп жіберу, дұрыс жолға салу нағыз садақа. Исламда садақа бергеннен әрі қайырымды іс қылғаннан пенденің күнәсі кемиді дейді» (Бегманов, 2010: 330). Келепжіп қолөнер. Бояуға арнап әр жерінен буылған жіп шумағы, келептелген жіп. Тілашар этн. Мектеп табалдырығын аттаған балаларға арнап өткізілетін кішігірім той. Уықбау этн. Көшкенде уықты жинап, буып байлауға қолданылатын бау. Төстұғыр қолөнер. Төсті орнататын дөңгелек ағаш. Жоғарыда төстердің барлығы да қайың, қарағаш, емен сияқты қатты ағаштардан көлденең кесіліп алынған дөңбекке бекітіледі. Оны төстұғыр немесе мықан деп атайды. Тұлым этн. Қыз баланың кішкентай кезінде екі шекесіне өсіріп қойған шаш, тұлымшақ. Науысқы қолөнер. Темірдің тот басқан жерін ысқылап тазартатын құрал. Құйысқандық малшар. Өгіздің жүгі мойнына жылжып кетпеу үшін оның санынан айналдырып тартып байлайтын жіп. Бұршік таға малшар. Жер басар алдыңғы жағында үш-төрт жерден шығыңқы етіп келтірген шошақтары бар таға түрі. Жамбаскиіз этн. Үйші шебердің киізден төртбұрышты етіп жасалған жабдығы. Етқошан тамақ. Қой құйрығы мен бұрыш, камыр және т.б. салынған қарынды пешке қойып пісіретін тамақ. Жасаусырмақ этн. Ұзатылатын қыз жасауына арнап жасалған сырмақ түрлері. Жүгіну қолөнер. Арқан тіндерін ширатып тартып жатқанда босатып алған тінде түйін, буылтық пайда болу, арқан тінінің түйінделуі. Кемпіржиек қолөнер. Жиекжіпті белінен бастырып тігуден пайда болатын жиек түрі. Қорып қалу қолөнер. Оттықтағы көмірдің түйін-түйін болып қалуы.

Сары жаулық – ертеректе қаралы әйел тағатын орамал. «Сары жаулық – күн төлеуде символдық және ғұрыптық бұйым ретінде жүрді. Күйеуі кісі қолынан өліп, жесір қалған әйелге күйеуін өлтірген адамның туыстары сары түсті жаулық беріп жіберу арқылы күн төлейтіндігін білдірген екен. Марқұмның әйелі бір жыл бойы осы сары жаулықты тартып жүруі қажет болған. Яғни, қаралы әйел екенін, кісі қолынан өлген күйеуінің құны төленетінін білдірген жосын болып саналады» (Әлімбаев, 2017: 708).

Сары сөзінің бір мағынасы қайғы-мұңға байланысты екендігі белгілі. Жырларда кездесетін сары бауыр ат тіркесі де осыған байланысты қалыптасқан:

«Мінгенде атың сары бауыр,

Шақырайын бір дәуір.

Шақырғанда келмесең,

Үстіңнен басар құм ауыр – дегендегі сары бауыр ат жай ұйқас үшін алынбаған, сол адамның асына сойылған тұл аты болуы керек. Ежелгі наным-сенім бойынша өлген адамның атын сойып, бірге жерлейтіні оны ол дүниеде де мініп жүрсін деген түсініктен туған. Ал, ол өлеңде, осы түсінік бейнелі көрініс тапқан» дейді зерттеуші (Ісләмжанұлы, 2007: 288).

Ауылдың үлкен кісілері мен ер-азаматтары жоқ кезде жұмыстан қолдары босаған әйелдердің бір үйге жиналып шай ішіп, әңгімелесуі, сырласуына мүмкіндік беретін басқосу абысыншай деп аталған. Бұндай салт-дәстүрдің қазіргі тұрмыста қайта қолданылып жүргенін атап өткеніміз жөн.

Мерзімді басылымдардан басқа жаңғырған атауларды Ж.Бабалықұлының малға, саятшылыққа байланысты еңбектерінен, Ә.Ысқақұлының сырмаққа С.Кенжеахметұлының салт-дәстүрге, Д.Шоқпарұлының қолөнерге, Б.Қалиұлының өсімдіктерге, сондай-ақ біраз зерттеушілердің дінтануға байланысты кітаптарынан, әр кезде шығып жатқан ғылыми-танымдық жинақтардан кездестіруге болады.

Әсіресе академик Ә. Қайдардың 3 томдық этнолингвистикалық сөздігін, Ш.Ш.Уәлиханов атындағы Тарих және этнология институтының «Қазақ халқының дәстүрлері мен әдет-ғұрыптары (құрастырған С. Әжіғали) 2 томдық мақалалар жинағы мен Қазақстан Республикасы Мемлекеттік орталық музейінде Нұрсан Әлімбайдың ғылыми жобасымен, жетекшілігімен дайындалып, екінші рет жарияланған 5 томдық «Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. Энциклопедия» атты еңбектердің орны ерекше.

Аталмыш энциклопедияның өзіндік ерекшелігі мен орны туралы былай деп атап көрсетеді. «Авторлар жойылуға айналған немесе мағынасы солғындаған, тіпті мүлдем сөнген, көптеген этномәдени категорияларды, ұғымдар мен атауларды ғылыми айналымға енгізіп отыр. Мұндай орны толмас өкінішті жайттың бірден-бір себебі – бодандыққа түскен халықтың экономикалық қана емес, рухани және тіл жағынан да бодан болады екен. Ана тілі мен ұлтының рухына, мәдениетіне немқұрайлы қарамайтын кім де болса, пұшайман етер сұмдық кеп – осындай рух пен тіл бодандығынан арылу «аяқ-қолыңды бос қойса да» қиынның қиыны екен. Әсіресе қазіргі жаһандану жағдайында...

Ғалымдардың міндеті – ана тіліміздің табиғи-тарихи қолданыстық кеңістігін қалпына келтіру арқылы оның әлеуметтік-рухани қауқарын қамтамасыз етуге қал-қадері жеткенше атсалысу. Әйтпесе, ана тілімізді «өлтіріп» алуымыз әбден кәдік. Өйткені тіл де күтіп-баптауды қажет ететін өзінше тірі организм ғой. Данышпан Ахмет Байтұрсынұлы «сөзі жоғалған жұрттың өзі де жоғалады» деген ой-тұжырымы аталмыш жайттың сұмдық салдарын айғақтап тұрған жоқ па? Авторлардың энциклопедияны жасақтау барысындағы сан-салалы ізденістері қазақ тілінің қазақтың және басқа жұрттардың талай зиялылары атаған шексіз байлығын, һәм ерекше қолданыстық икемділігін тағы да дәлелдегендей» (Энциклопедия, 2017: 11).

Бұл энциклопедиядан он бес томдық «Қазақ әдеби тілінің сөздігіне» еңбеген сөздер мен сөз тіркестерін көптеп кездестіруге болады. Мәселен Б.Сарыбаев, Т.Әсемқұлов, Ө.Жәнібеков, С.Медеубекұлы, Д.Шоқпарұлы, Б.Орынбеков, Ж.Нәжімеденов, Н.Мұқатайұлы, З.С.Жәкішева және т.б. зерттеушілердің еңбектерін жинақтай отырып жазылған «Домбыра» атты мақалада домбыраның құрылысына байланысты қауыншақ, тұғылда, ернеулік, мойынбет, алқымдық, ал пернелеріне байланысты бас перне, оғыз перне, ұры перне, тылсым перне, қосақ перне, еңірелі перне, тірек перне, жетім перне, жуас перне, азат перне, бел перне, көсем перне, мұңлық перне, таса перне, даңғыл перне, асу перне, шешен перне, еңіс перне, етек перне, үйсіз перне, төс перне, шағыршақ перне, алқым перне, құсар перне, шықшыт перне, тамақ перне, оймауыт перне және т.б. 40-тан астам атаулары келтіріліп, олардың тілдік мағынасы мен мәдени семантикасы ашылған (Энциклопедия, 2017: 75).

Энциклопедиядан қазіргі көпшілікке беймәлім тектұрмас, атөткел, қанды қақпан, тұздық ашу, бестеше, қоздату, көкқасқа, қостаушы, шабысты жүріс, атсырау ережесі, белбасар және т.б. тоғызқұмалақ ойынына байланысты бір топ атаулар қамтылғанын ұшырастыруға болады. Бұндай сөздер оқырманның таным көкжиегін кеңейте түсері сөзсіз.

Жаңғырған атаулардың ұлттық тәрбие беруде маңызы зор. Өйткені жаңғырған атаулар танымдық қызметі, дамытушылық қызметі, әлеуметтендіру қызметі, коммуникативтік қызметі, ақпараттық қызметі, бірегейлендіру қызметі, сабақтастық қызметі тәрізді атқарушылық қызметтері арқылы адамның «рухани өмірін қалыптастыру механизмі болып табылады» (Сәрсенбекова, 2010: 95).

Жаңғырған атауларды жинау, зерттеу, оларды жүйеге келтіріп лексикографиялау бүгінгі күнде өзекті болып отыр. Тіл ғылымының алдында заман рухына сәйкес тілдік құндылықтарды сақтау, тілдегі құндылықтарды жаңғырту, тілдегі құндылықтарды бейімдеу, тілдегі құндылықтарды қайта түлету және т.б. аса маңызды міндеттер туындады. Бұндай міндеттерді жолға қойып, шешу жолдарының бірі лексикографиялаумен байланысты.

Ұлттық тіліміздің сөз байлығы лексикографияланып, көптомдық «Қазақ әдеби тілінің түсіндірме сөздігі», біртөмдік «Қазақ сөздігі» жарық көрді. Бұларды ғылыми-мәдени жұртшылық оны тіл білімінің зор жетістігі деп бағалап, әсіресе ондағы этномәдени семантиканың берілуін аталмыш сөздіктердің жаңалығы деп айрықша атап көрсетті.

Алайда біріншіден, халық тілі қойнауында лексика-фразеологиялық бірліктердің әлі де болса сөздіктерде қамтылмай жатқаны, екіншіден лексика-фразеологиялық бірліктерде әлі де болса этномәдени семантикаға тән компоненттерді аша түсу керектігі айқын байқалады. Осы аталған мақсат-міндеттерге байланысты:

- колөнер, салт-дәстүрлер, табиғат объектілері, діни лексика, қоғамдық қатынастар саласындағы атаулар жинақталды;
- жинақталған лексика-фразеологиялық бірліктер тезаурус формасында лексикографияланып, электронды нұсқасы жасалды.

Зерттеу тақырыбының нәтижесінде алынған өнімдер «Қазақ тілінің ұлттық корпусына» пайдаланылады; сондай-ақ алдағы уақытта қолға алынатын «Үлкен біртөмдік қазақ сөздігіне», көптомдық түсіндірме сөздікке, екі тілді аударма сөздіктерге, лингвомәдени сөздікке қажетті лексикографиялық материал болып табылады.

8.7 Этномәдени бірліктердің корпуста берілу жолдары

Қазіргі кезде жаһандануға байланысты жаңа ақпараттарды игеру мақсатында орыс тіл білімінде немесе шетел тіл білімінде үлкен жетістіктерге жетіп отырған тілдік корпустар бар. Шетелдік ғалым Э.Финеган корпустық лингвистикаға табиғи тілдің қолданысын зерттеуге бағытталған корпусты құрастыру мен қолдануды қажет ететін қызмет түрі деген анықтама береді [222]. Заманауи Қазақстанның да рухани-танымдық жаңа деңгейі Қазақ тілінің ұлттық корпус жасау қажеттілігін көрсетіп отыр. Өйткені, Ұлттық корпус-

тарды құру мәселесі шетел тіл білімінде ХХ ғ. 60-жылдары басталса, орыс тіл білімінде ХХ ғ. 80-жылдары, ал қазақ тіл білімінде осы соңғы жылдарда ғана қолға алынды. Қазақ тілінің корпустық лингвистика мәселесін профессор А.Қ. Жұбанов көтерді. Ғалымның пікірінше, корпустық лингвистика қазақ тіл білімінің де ерекше саласы ретінде қалыптасатын болса, қазақ тілшілеріне көлемді тәжірбиелік материалдарды пайдалануға, қажетті деген тілдік деректерді тез тауып алуға және оларға тиісті деген өңдеулер жүргізуге мүмкіндік тудырады. Осының бәрі қазақ тіліне қатысты зерттеулердің шынайылыққа (ақиқаттыққа) жетудің эмпирикалық тәсілдеріне жаңаша көзқараспен қарауға және ғылыми айналым аясына аса маңызды тілдік материалдарды енгізуге жағдай жасайды [223, 4].

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

Лингвистер мәтіндерді арнайы бағдарламалардың көмегімен өңдейді. Бұл, әрине, автоматты түрде жасалады. Корпус неғұрлым үлкен болса, соғұрлым жақсы. 100 миллион сөзқолданыстан тұратын корпус ұзақ уақыт бойы көлемі мен саны жағынан үлкен корпус болып саналды. Ал қазір миллиардтаған базадан тұратын корпус бар. Осылайша, өңделген мәтіндер, аннотацияланған мәтіндер корпусы құрайды. Корпус бізге зерттеу жұмысын жүргізуде мысалдарды тез әрі жан-жақты іздеу үшін қажет. Бұрындары материалдарды жинауға айлар емес, жылдар да кететін болса, қазір Корпус материалды қолдап жинау жұмысын өте қысқартып, уақытты үнемдеуге септігін тигізуде.

Тілдік корпусарды құрастырудың қиындығын айқындап, саралап берген профессор А.Қ. Жұбанов ҚТҰК-ны құрастыруды орталықтандыру, мемлекеттік дәрежеде барлығы бірдей жұмыла кірісу керектігі туралы 2014 жылғы «Ана тілі» газетінің бетінде осы мәселеге қатысты бағдарламалық мақала жариялады. Ғалым ол туралы былай дейді: «Ұлттық тіл мәтіндерінің компьютерлік корпусын құру жобасы бір ғана ғылыми ұйымның шешетін мәселесі емес және ол зерттеу жұмысы 3-5 жылда аяқтала қояды деуге де болмайды. Себебі, бұл аса күрделі және оның нәтижелері әлемдік дәрежедегі аса маңызды ғылыми жұмыс болып саналады. Зерттеу жұмысының мақсатына сай орындалатын міндеттері сала-салаға, кезең-кезеңге бөлініп, тек қана ғалымдар қауымдастығын құру арқылы ғана ауқымды нәтижеге жетуге болары сөзсіз.

Қазақ тілінің ұлттық корпустарын жасау – көп болып жұмылып атқаратын күрделі іс. Сондықтан Қазақстанның әр жерінде бірінен-бірі дербес атқарылып жатқан корпус жасау ісін орталықтандыру керек. Ол үшін Қазақстанның әр жерінде ғылыми-педагогикалық қызмет атқарып жүрген ғалымдардан арнайы лингвистикалық топ құру керек. Өйткені тілдік бірліктерді модельдеу – өте күрделі мәселе Сондай-ақ тілдік корпустарды құрастырудың өзі, ең алдымен, лингвистикалық белгіленім талдамасын жасауға кеп тіреледі. Осы кезге дейін жасалып жатқан корпус жасау тәжірибесіндегі нәтижелерді бір орталыққа жинақтау керек. Кітап, газет-журнал шығаратын баспалармен шартқа отырып, мәтіндердің электронды нұсқаларын алу қажет» – дейді [224, 9].

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

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

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

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

Қазіргі кезде қазақ тіл білімінде тілдік корпустарды құрастыруға қатысты қазақ тілінің тілдік деңгейінің зерттеу нысанына сәйкес лингвистикалық белгіленімдер, морфологиялық деңгейде белгіленім қою – аннотациялау жұмыстары жүргізілуде. Орыс тіл білімінде лингвистикалық ақпаратты «разметка» деп атайды. Ал, бұл терминді профессор А.Қ. Жұбанов қазақ тілінде «белгіленім» немесе «белгі-код» терминдерімен атауды ұсынды. Тілдік корпустарды жасауда осы терминдер белсенді қолданылып жүр. Корпустарда этномәдени бірліктердің белгіленімдерінің сипатына қарай оларға (тегтар) пайдаланылады. Мұндағы белгі-код «тег» сөзі (ағыл. tag – ярлық). Аннотациялау дегеніміз – қазақ әдеби тілінен сұрыпталып алынған таңдама мәтіндердегі сөзформаларға лингвистикалық мәліметтер беріп, ақпараттандыру. Лингвистикалық белгі-код қою арқылы тілдік мәліметтерді бейнелесе, белгілер семантикалық (лексика-семантикалық), морфологиялық, синтаксистік сипатта болады. Ұлттық корпуста сөз табына, септелуіне, тегіне, жіктелуіне, шырай түріне, жанды-жансыздығына қарай оларға морфологиялық, синтаксистік белгіленімдер қойылады. Ал, семантикалық талдаулар арқылы мәтін ішінде қажетті белгіленімдер қойылады. Корпус жасағанда белгіленімдердің барлығын бір уақытта енгізу қиындық туғызды. Корпустарда лингвистикалық белгіленімдер кезең-

кезеңімен жүзеге асырылып отырады. Корпуста этномәдени белгіленімдер қою шартты түрде арнайы тәгтерді «тіркеп» жазу арқылы белгіленетін болады.

Бүгінгі ақпараттық технология заманында кез келген ғылым саласын цифрландыру мәселесі гуманитарлық ғылым салаларына да қажеттілік туғызды. Өйткені қазір барлық қоғамдық қызмет түрлері мен оқу жүйесі компьютерлік техника арқылы орындалуда. Ғылым дамыған сайын кез келген қызмет түрін жеңілдету, оңтайландыру қажеттік тудырып отыр. Қазақ тілінің ұлттық корпусын жасау да лингвистикалық зерттеулер жүргізудің мүмкіндігін арттыру, мемлекеттік тілді корпустық ресурстар арқылы жедел меңгерту мақсатын көздейді. Қазақ тілінің ұлттық корпусы, негізінен, этнос тіліндегі материалдық және танымдық мәдени ақпараттарды жеткізуші қызметті атқаратын құрал, сондықтан Ұлттық корпуста этномәдени бірліктерді енгізудің маңызы зор.

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

ҚТҰК-да этномәдени бірліктерді кодтаудың технологиясын жасау және этнобелгіленімдерді енгізу жұмыстарын да қарастырудың маңызы зор.

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

Ш. Уәлиханов, Қ. Жұбанов, Ә. Марғұлан, М. Әуезов өз еңбектерінде тіл мен мәдениеттің бір-бірімен тығыз байланыстылығына мән бергені белгілі. Мысалы, Қ. Жұбанов тіл арқылы халықтың тағылымдық мұраты мәдени әрі тілдік мұра ретінде сақталатынын және этностың әр дәуірдегі ұрпақ өкілдеріне тіл мен мәдениетті табыстыратын жаңа саланың дүниеге келетіні туралы сол кезеңдердің өзінде-ақ болжам айтқан. Ғалымның «халықтардың тұрмыс қалпы түрлі-түрлі болған соң, олардың әр затқа қоятын аттары да түрлі-түрлі. Тіл ұзақ заман жасалады, бірден өзгермейді, оның өзгерісі тұрмысқа байланысты» – деген пікірі ұлт мәдениетінің ұрпақ жалғастығын сипаттаса, ал, мәдениет жай өмірдің туындысы ғана емес, белгілі бір қоғамда өмір сүрген рухани әрі материалдық қоры бар іргелі елдің өмір тіршілігінің қажетті туындысы деп

қарауын: «Мәдениет – қазанның күйесінше жұғатын нәрсе емес, мәдениеттің белгілі бір сатысы көкसेгендей жағдай тумаған болса, жоғары мәдениеттің өзі келіп ауызға түспейді» деген пікілерінен көруімізге болады [225, 7].

Тіл мен мәдениеттің байланысын лингвомәдениеттану саласы құрайды. Лингвомәдениеттанушы ғалым В.И. Телия лингвомәдениеттану саласының басты мақсатын былай түсіндіреді: «Тіл мен мәдениеттің өзара қатынасын қазіргі заман мәдени ұлттық сана-сезім тұрғысынан қарастырып, оның тілдегі көрінісін зертеу мен сипаттау» [226, 28].

Мәдениеттің негізгі арқауы тілде сақтала отырып, сол арқылы ұрпақтан-ұрпаққа жетіп отырады. «Тіл – өмір шындығы. Сол шындықты уақыт қалдырған қатпарлардан аршып ала білсек, шежіредей шыр шертіп тұрады. Өмірде болған нәрсе тілде де із тастамай кетпейді. Сондай іздің сорабын асыл мұрамыздың тілінен кездестіреміз» [227, 13].

Ал «белгіленім» – корпустан іздеген сөзіңіз туралы толық ақпараттарды көрсететін термин. Қазақ тілінің ұлттық корпус базасында бес түрлі белгіленім бар: морфологиялық, сөзжасамдық, семантикалық, лексикалық, фонетика-фонологиялық белгіленімдер. Біздің алға қойған мақсатымыз – қазақ тілінің ұлттық корпусына этномәдени бірліктерді енгізу. Оларды түрлеріне қарай топтастырып, өңдеп, одан әрі жетілдіру этномәдени бірліктерді корпуста енгізіп, оларға шартты белгілер қою арқылы жүзеге асырғаннан кейін мына міндеттер орындалуы тиіс:

1. ҚТҰК-ға этномәдени бірліктер қорын беру үшін Excell бағдарламасында бастапқы мәтіндер базасы жиналады;

2. Excell бағдарламасына жиналған этномәдени бірліктердің мәтіндерін негізгі базаға енгізу үшін өңделеді;

3. Excell бағдарламасына жиналған этномәдени бірліктер негізгі базаға енгізіледі және мәдени-семантикалық корпуспен жалғанады;

4. Корпуста этномәдени бірліктерге белгіленім қойылады. Мысалы; (этно) белгіленімі;

5. Корпустардың мәдени-семантикасы бойынша іздеу (поиск) жасалады;

6. Әрбір белгіленімдер бойынша іздеу жұмысы жетілдіріледі.

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

1. Этномәдени бірліктерді корпуста жинаған кезде, ең алдымен, саба туралы анықтамалық ақпарат жиналады;

2. Саба – кептірілген шикі теріні әдемі құрамға салып, құм кептеп қатырып, содан кейін ысқа салып төрт бойдан құрастырылған тері ыдыс;

3. Сабаның ең басты белгісі туралы ақпараттар жинақталады: сабасыз іркіт жиналып, қымыз құйылмайды;

4. Сабаға (этно) белгіленімі қойылады;

5. Сабаның түрлері туралы ақпарат жиналады. Саба – жылқы терісінен жасалып, терінің түгі жидітіліп, терісі сыпырылып, көкетін алып тазартылды. Дайын болған тері талдың қабығымен боялады. Ол шамамен 100 литр болатындай түбі кең, аузы тарлау болатындай етіп пішіліп, қосылған жерлері та-

рамыс немесе шуда жіппен жөрмелеп тігеледі. Сабаның ішінде қымыз жарық түскенде түбі көрініп тұрады. Ысталған ыдыс берік болады, құрт-құмырсқа түспейді, әрі ыстың дәмі мен иісі шауып, оның иісін арттыра түседі. Сабаны іркіт жинап, қымыз пісу үшін пайдаланылады. Сабаның бұлқын, қарақұла сияқты түрлері бар:

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

Қарақұла – сабаның жүні мен қылы жидітілген теріден жасалады. Жылқының жал тұсынан сабаның ортаңғы бөлігіне тігінен келтіріп тігетін болған. Сабаның қарақұласы көзге әдемі көрінетін болған. Сабаны жерге қоймаған. Оның астына пішініне лайықтап сәкі (орындық) жасайды немесе оны саба орындық деп атайды. Сабаның астыңғы жағына күнтулақ немесе киіз салып пайдаланады. Сабаны үйдің кіре берісіне қойып, іркіт жинап отырған. Сондықтан да қымыз сабасы іркіт сабасынан жоғары тұрады. Сабаның аузын керегенің басына байлап қояды. Қымызды пісу үшін піспек пайдаланылады [210, 115].

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

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

Корпустарда заттық атаулардың олардың қолданысына қарай мағыналық динамикасы да тілде көрініс табады. Мысалы: тегене – ағаштан жасалған шұңғыл тік жиек ыдыс. Тегене қатты ағаштан тұтас шабылып, ойып жасалады. Тегененің шұңғыл жасалған түрі де кездеседі. Тегененің екі құлағы болады. Құлағы осы ағаштан шауып жасалатын болса, кейде темірден немесе жезден де жасалады. Тегене – қымыз, шұбат сияқты түрлі сусындарды құюға арналған қолайлы ыдыс. Тегененің сыртын ағаштан оюлап безендіреді. Қымызға арналған тегенелерді сүйектен асыл тастармен әшекейлеп те қолданған. Оюдан ойылып жасалған қақпағы болған. Өйткені, тегененің көлемі үлкен болғандықтан, күнделікті өмірде көп қолданылмайды, тек той жиындарда ғана қолданылады. В.Плотниковтың айтуы бойынша: тегене көлемі *5-15 литрге* дейін жеткен. Ағашы қаңсып шытынап кетпес үшін, күн көзіне тигізбей оны сүртіп, майлап отырған [210, 385].

Осындан қазақ халқының тұрмыстық бұйымдарды асқан ұқыптылықпен қолданғаны көрініп тұр. Дәстүрлі ортада тегенелердің оюларына қарап, үйдің әл-ауқатын байқайтын болған.

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

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

Үнемі өзімен бірге малшылар сусын құйып алып жүргендіктен, жанторсық деп аталады. Мүйізді торсықтың сыртына түрлі ою-өрнектер салынады. Оның оюланып, әдеміленіп жасалған түрі кеукерін деп аталады. Пішіліп алынған терілер бір-бірімен беттестіріліп қойылып, олардың шеттері тарамыспен немесе шуда жіппен тігілген. Оның екі бүйіріне батырғы (батырма) құрал ұшымен өрнек салынады. Мойын тұсынан кішкене «құлақ» шығарыла пішіліп, оған тұйықтап темір шегелеп, дөңгелек етіп кішкене ғана тесік тесіліп, оған бүтін бір қайыс таспа тағылған. Ол таспаны өріп бау таққан. Торсық жасалып біткеннен кейін кептіріліп, ол кепкеннен кейін оның ішіндегі құмы төгіліп, жуып тазартылады. Осыдан кейін торсық кептіріліп, маймен майланып, тобылғымен ысталып, сусын қоюға қолданылған [210, 454].

Сүйретпе – жылқының бас терісін шикі күйінде жібітіп, түтігін алып, бітеулей тігіп, ысқа 7-8 күн қойып дайындалған қымыз сақтайтын ыдыс, сүттен жасалған сусын түрлерін сақтауға, оларды тасымалдауға арналған теріден жасалған ыдыс. Ол – көлемі жағынан сабадан кіші ыдыс. Сүйретпе – ұзынша, түбі сопақтау келген мойыны қысқаша. Сүйретпенің негізгі қызметі – сабаны босатып алу үшін сабадағы қымызды қою. Қымызды сүйретпеді уақытша сақтауға, онымен тасымалдауға болады. М.Ж.Көпейұлының айтуынша: Абылай Жарылғаптың аулына келгенде Маңқай бәйбіше жүктің астында сүйретпеді байдың сыбағасы осы түнемел тұр еді, – дейді [210, 358].

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

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

Тіл корпусарын тіл білімінің барлық саласында қолдануға болады. Нақтырақ айтсақ, әртүрлі сөздіктерді құрастыруда (жиілік, түсіндірме, аймақтық, фразеологиялық, энтографиялық), оқулықтар құрастыруда, аудармалар жасауда қолданылады.

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

Ең бастысы ешбір корпуста этномәдени бірліктер берілмеген. Осымен байланысты болашақ зерттеу жұмысымызда этномәдени бірліктерді ҚТҰК-да қалай беру керек деген мәселе өзектеледі. Зерттеу нәтижесінде ҚТҰК-ға этномәдени белгіленімдер енгізу жоспарланып отыр.

Одан бұрын қазақ тіліндегі лексикографиялық және корпустық зерттеулердің теориялық мәселелері толықтырылып, этномәдени бірліктерді лексикографиялаудың әдіс-тәсілдері айқындалу қажет.

8.8 Лингвомәдени білімдер жүйесін ұлттық корпусқа енгізуде көне атауларды зерттеу

Материалдық мәдениет те, рухани мәдениет те тілде көрініс тауып отырады. Мәдениеттің қай-қайсы болмасын, ұлттық мәнділікке жатады. Оның ұлттық сипаты әр халықтың ақиқат дүниені тілі арқылы өзіндік көзбен көруіне байланысты болады. Тіл – қарым-қатынас, танымның құралы ғана емес, тіл – ұлттың мәдени коды. Тілдің жәрдемісіз мәдениетті оқып-игеру, байыптау мүмкін емес. Тіл арқылы ұлттың қазіргі кездегі менталитетін, байырғы кездегі тілдік ұжымның дүниеге, қоғамға, өзіне деген көзқарасын, сана-сезімін айқындауға болады.

Қазіргі кезде көпшілікке қолжетімді Ұлттық корпус дегеніміз – белгілі бір тілдің табиғатын толық бейнелей алатын, тілдегі барлық стиль мәтіндерін қамтитын, көлемді, теңгерімді (репрезентативті) корпус [12, 48]. ҚТҰК-да қазақ тілінің 30 миллион сөзқолданыстан тұратын электронды мәтіндер қоры жинақталған (<https://qazcorpus.kz/>). ҚТҰК-ға енгізілген мәтіндер қорына морфологиялық, лексикалық, семантикалық, фонетика-фонологиялық белгіленімдер қойылады. Қазіргі кезде ҚТҰК құрамында негізгі корпуспен қатар диалектілік, тарихи, ауызша, параллель, мәдени ішкорпустар да қатар жасалып жатыр. Ішкорпустар негізгі корпустан мәтіндік базасы, белгіленім, метабелгіленім, іздеу жүйесі бойынша ерекшеленеді. Аталмыш мақала ҚТҰК-ның құрамына енетін мәдени ішкорпусқа лингвомәдени бірліктер жинау мәселесіне арналып отыр.

Қазіргі ақпараттық қоғамда мәдени құндылықтарды келер ұрпаққа жеткізіп, сақтаудың негізгі көзі – мәдени ұғымдарды Ұлттық корпустар базасына енгізу. Корпус – ұлт тіліндегі барлық стильдік қабатты қамтитын мәтіндердің бағдарламалық тұрғыдан қамтамасыз етілген электрондық жинағы, жинақтаушы банк. Корпустың қай түрінде болмасын, ерекшеліктеріне қарай мәтіндердің жинақталуы басты шарт болып табылады. Сонымен қатар Ұлттық корпустардағы екінші басты шарт – лингвистикалық белгіленімдермен жабдықталуы [228, 472-473]. Осы ретте морфологиялық белгіленім, просодикалық белгіленім, семантикалық белгіленім т.б. тіл деңгейлеріне байланысты белгіленім түрлері беріледі. Кез келген сөздің қандай мағынаны білдіретіндігі туралы ақпарат та пайдаланушы үшін қажетті сұраныс тудырады. Сондықтан Ұлттық корпустардан ізделген сөздің лексикалық мағынасымен бірге, мәдени бірліктердің мәдени семантикасы туралы да ақпарат беру, ең алдымен, ұлт тілұстарманы үшін маңызды болып табылады.

Лингвомәдени бірліктер мен олар туралы мәдени білімдер жүйесі және мәдени мәтіндерді жинақтаушы банк, ол – Қазақ тілінің мәдени-репрезентативті ішкорпусы. Бұл ішкорпус түрі – қазақ тілінің кең ауқымды мәдени-лингвистикалық қоры болғандықтан, оған енетін мәтіндер мен мәдени бірліктер тақырыптары жан-жақты болып келеді.

Мәдени-семантикалық белгіленім – сөздердің мағыналық құрылымында кодталған ұлт мәдениеті туралы білімдер жүйесі. Мәдени білімдер жүйесін ішкі формасында сақтаушы бірлік, ол – сөз, яғни мәдени бірліктер болып табылады. Осы тұрғыдан алғанда, әдеттегі түсіндірме сөздіктер сөздерді тіркеуші және сөздердің лексикалық мағынасы туралы ақпарат беруші құрал болса, мәдени сөздіктер сөздердің лексикалық мағынасын емес, сөздердің шығу уәжімен байланысты ұлттық салт-дәстүр, тұрмыс-салтпен байланысты мәдени семантиканы да түсіндіреді. Ал мәдени сөздіктер корпустағы этнолингвистикалық белгіленімдерді (мәдени-семантикалық) жасауға негіз болады.

Мәдени сөздіктер жасау ісі бір күннің ішінде жасай салатын жұмыс емес, ол ұлттың өмір сүру кезеңдерімен бірге жасасып, ұрпақтан-ұрпаққа тарап, қазірге дейін жетіп келе жатқан ұлттық құндылықтарымызды сараптау, тіркеу, жүйелеу, зерттеу, тағы да басқа ізденістер арқылы жүзеге асады. Мәдени сөздіктерді жасаудың дереккөздері:

- халықтың салт-дәстүрі, ұғымдары туралы білімдер;
- этнографиялық еңбектер;
- фольклорлық еңбектер;
- тарихи еңбектер;
- мәдени бірліктер туралы жазылған зерттеулер (диссертациялар, мақалалар);
- мәдени бірліктер семантикасын ашу т.б. [229].

Аталған дереккөздердің ішінде мәдени бірліктерге зерттеу жүргізу қажеттігі туындады. Қазақ тілінің лингвомәдени ішкорпусына енгізілетін көне атаулар лексикалық байлығымыздың ерекше бір саласын құрайтын, бірыңғай тақырыптық жүйе болғандықтан, олардың мәдени семантикасын ашу үшін атақ-лауазымға қатысты тілдік деректерді ғылыми тұрғыдан жүйелей отырып, оларды дәуірлік, мағыналық топтарға жіктеп, олардың этнолингвистикалық, уәжділік және ішінара этимологиялық табиғатын тілдің қалыптасқан заңдылықтары аясында саралап көрсетудің маңызы ерекше.

Бүгінгі қазақ тіл білімінде қалыптасқан, өзіндік ұстаным-үрдістері бар этнолингвистикалық зерттеулерде (Ә.Қайдар, М.М.Копыленко, Е.Жанпейісов, Ж.Манкеева, Р.Шойбеков, С.Сәтенова, Г.Смағұлова және т.б.) этностың салт-санасы мен өмір-тіршілігінде қалыптасқан ұғым-түсініктердің, атаулардың ұлттық сипатын, жалпы қазақ этносының ұлттық бейнесін, өзіндік болмысын толық та түбегейлі түрде тек оның байлығы – «тіл әлемі» арқылы ғана танып-білуге болатындығы дәлелденіп отыр.

Лингвомәдени жүйеге жататын тілдік бірліктердің ең басты ерекшелігі – өз бастауын сонау көне дәуірлерден алып, үнемі даму үстінде қалыптасып, даму үдерісінде толығып, өзгеріп, жаңарып отыруында. Сондықтан да, лексиканың бұл саласы – қазақ этносының сатылы дамуының, қоғамдық өзгерісінің

ру, тайпа, ұлыс, халық, ұлт болып қалыптасу кезеңдерінің бірден-бір куәгері, халықтың өзімен де, тілімен де бірге дамып, біте қайнасып келе жатқан көне мұрасы. Осы тұрғыдан алып қарағанда жалпы, түркі тілдеріне тән және солардың бірі – қазақ тіліндегі толық және жартылай ортақ атақ-лауазым атауларын 5 дәуірге бөліп қарауға болады. Осындай дәуірлеуге байланысты оларды мынандай тақырыптарға бөліп қарастырамыз: 1. ғұн дәуіріндегі атаулар (б.з.д. V- б.з. V ғ. дейін); 2. көнетүркі дәуіріндегі атаулар (VI-XII ғ.ғ.); 3. ортатүркі дәуіріндегі атаулар (XIII-XIV ғ.ғ.); 4. жаңатүркі дәуіріндегі атаулар (XV-XIX ғ.ғ.); 5. Қазақ хандығының отарлануы дәуіріндегі атаулар (XIX ғ.).

Бұл хронологиялық межелер біздің шартты түрде алып отырған дәуірлерімізге жатады да, түркі әлемі тарихында оны бұлайша жіктеу байқалмайды.

Ғұн дәуіріндегі атаулар

Тарихшылардың зерттеуі бойынша, «Түркі елінің ежелгі тарихы біздің заманымыздан бұрынғы III-ғасырдан, ертедегі ғұндар империясының құрылуы кезеңінен басталады да, б.з. IX-ғасырын қамтиды. Біз жарияланған қытайша қолжазбаларды қарастырып, ғұн дәуірінде қолданылған атаулардың жалпы сипатына тоқталып өтуді жөн көрдік. Сол қолжазбалардың біріне – «Сиунну» кітабынан аударып, құрастырылған «Таным тармақтары: Қытай тарихшыларының көне тарихымыз жөніндегі пайымдаулары» атты еңбек, екіншісіне – «Ғұн тарихының жылнамасы» атты қолжазбалар жатады [230]. Түркітанушы ғалымдардың пікірі бойынша ғұн туралы қытай естеліктерінде кездесетін лауазым атаулары саны өте аз және қытайша жазылуы олардың нұсқасын дәл бере алмайды дегенді себеп етті. Осының салдарынан ғұндар мен түркілердің генеологиялық ортақтығы көп сөз болмай келгені белгілі. Алайда қытай қолжазбаларының табиғаты, әсіресе кейінгі кезде кеңірек ашылып, ғылыми айналымда түркі тілдеріне аударылған тарихи деректердің көбейе түсуі оларды зерттеушілердің бұл құбылысқа ерекше назар аударуына себепкер болды.

Осы ретте зерттеуші Қ.Ғабитханұлының ғылыми мақаласындағы келтірген кейбір дәйектемелеріне тоқтала кеткен орынды. Зерттеушінің пікірінше, кезіндегі орыс миссионері Бичурин (1851) мен француз ғалымы Э.Шаванның (1880-1970) аудармаларындағы ономастикалық атаулар сол күйінде қабылданауда. Қытай жылнамаларындағы ономастикалық атаулардың біразының қалай оқылатындығы бүгінде белгілі болса да, тарихшыларымыз ғылымның лингвистикалық жағын ескерместен, тарихына ғана мән беруде. Ономастикалық атауларды қолдану бір жүйеге түспеген. Мәселен, ғылыми әдебиеттерде ғұндардың алғашқы тәңірқұттарының бірінің аты қазіргі оқылуы бойынша – *Mao dun*, оны Бодұн, Мөде, Моде, Маудун, Мете, Модо, Модэ, Мао-дунь т.б. деп, Үйсін елінің басшысы қазіргі оқылуы бойынша – *Kun mo*, яғни Күнби, Күнмо, Күнми, Күнбек, Гүнмо т.б. түрінде жазылуда. Сол сияқты, тәңірқұт атауына келсек, оның қазіргі оқылуы *cheng li*, ежелгі оқылуы *dang lei*. Бұл мағыналық жағынан тәңірі сөзіне жақын. Қазіргі оқылуы *gu tu*, бұрын *go to* немесе *gua ta* деп оқылған. Бұл сөз дыбысталуы жағынан түркінің құт, құт(т)ы сөзіне жақын болғанымен, мағынасы жағынан кейбір түркі тілдерінде бала, сәби деген мағынада айтылатын гөдек (көдек) сөзіне жақын, яғни бұл қытайша бастау хаттағы *dang lei gua ta* тәңір көдегі сөзі тәңірден туған,

тәңір ұлы дегенді білдіреді деп түсіндіреді. Біз қалыптасқан дәстүр бойынша бұрынғыша тәңірқұт деп атадық. «Шүлен адам», «Шүлен беру», «Шүлен тарату» дегендегі «Шүлен» (*gi lian*) – ғұндарша тәңірдің басқаша аталуы. Ғұндар тәңірқұттың әйелін және күйеуге шыққан қыздарын осылай атаған, яғни көне түркідегі «қатұн» сөзі сияқты ғұн әйелінің мәртебесін білдіретін сөз. Бұл сөз дыбысталуы және әйел деген мағынасы жағынан «абысын-ажын» дегендегі «ажын» сөзіне жақын. Көне түркі тіліндегі атақ-лауазымға байланысты қытай жылнамаларындағы ономастикалық атаулардың қазірге салыстырып оқып анықталғандарының ішінде тегін, шад, елтебер, іркін, түртіп, бұйрық, Түмен қаған, Бөрі қаған, Үка қаған, Үка тегін, қаған, қатұн, ябғу, апа, елші, тұтұқ, санғұн (сардар) және т.б. бар [231, 94].

Бір артықшылығы, қолжазбаларда бұл атаулардың қытай тіліндегі аудармасы емес, сол өздерінің бастапқы нұсқасында, сол кездегі аталуынша және нақтылы мәтіндерде қолданылуы бойынша беріліп, басым бөлігінің мағынасы түсіндірілген. Бұл атаулардың түркі халықтарының тіліне тән екендігін бірден тануға болады: аға тәңірқұт, құт, аға, білге, хан, кіші, тете, бек, бас, жасауыл, уәли, ордабек, даяшы, бақауыл, күзетбасы, даяшыбасы, бас абыз, тұтқауыл, күнби, мыңбасы, жүзбасы, онбасы, бас жасауыл, төреші абыз, тәңірқұт, уәзірбасы, аламанбасы, аймақбасы, шабандоз, сол қол аламанбасы, қорғаушы тұтқауыл бек, сардарбасы, тілмаш, сол бас сақа, ішкі бек, басқақ, нақбек т.б.

Мұны дәлелдейтін екінші бір дерек – бұл атаулардың барлығы, әсіресе екі-үш сыңардан тұратын аналитикалық-синтетикалық сөзжасам моделі бойынша жасалған және ол үлгі түркі тілдерінің бір сипаты екендігі белгілі. Мысалы: құт аға, білге хан, кіші тете хан, бас жасауыл, бас уәли, оң білге хан, күзет басы, даяшы басы, бас абыз, мыңбасы, онбасы, жүзбасы, уәзірбасы, аламанбасы т.б.

Атаулардың біразы бас (главный, старший) мағынасындағы анықтауышпен тіркесіп жасалған. Мысалы: бас жасауыл – жасауылдардың мәртебесі биігі; бас уәли – уәлилердің басшысы; ордабасы – хан ордасын басқаратын лауазым; «Хан патшасы сарай күзетбасы Лиңгуанды атты әскер санғұндығына тағайындап, Юнжұнға жіберді» [230, 42]. Мағынасы бүгінде айқын атауларға, мысалы, хан, патша, бек, карауыл, абыз, уәли, даяшы, күзетбасы, онбасы, жүзбасы, мыңбасы, аламанбасы, аймақбасы, шабандоз, қорғаушы, сардарбасы, тілмаш т.б. осы тәрізді лауазымдарды көрсетуге болады. Ғұн тілінде жазылған тұтас жазба мәтіндері сақталмағандықтан, олардың қытайша аудармасын ғана көрсетуге болады. Олардың бірнешеуінің мағынасына тоқталып кетелік: төреші абыз – «хан патшалығы төреші абыз хан Ангоның ұсынысы бойынша құдаласуға келісім берді» деп жазылған; күнмо – бұрындары Үйсіннің ханын күнмо деп атайтын; бас абыз – хан патшалығы бас абыз Таочинді Дай бектігіне жіберіп, құда-андаласып татуласады; ордабек – хан патшасы осылардың бәріне ордабек – Лиеху деген атақ берді. Бұл атаулардың біразы қазіргі кезде түркі тектес халықтар түсінбейтін ғұн не қытай атауларын құрайтын болса, енді біразы түркілік атаулар екендігіне ешкім күмән келтіре алмайды.

«Көне дәуір күмбірі» атты қолжазбадағы Қытай тарихындағы қазаққа қатысты мынадай атақ-лауазым атауларының тізімі кездеседі (қазақ тіліне аударылып, басылған нұсқасынан): ұлыс басы – батыс түрік оғының ұлыс басы

тұтам сарайға опасыздық жасады (55 б.); түрік қағаны – жаз 25 сәуірде түрік қағаны Боғы Чор тағы да елші жіберіп, қыз айттыруды талап етті (55 б.); басқақ – патша жарлық түсіріп, Бесбалық басқағы Таң Жиахуй мен ақылшы нөкер жие Уанды қосын бастатып көмекке жіберді (55 б.); елші – Жаң Шиачун әр елге хат арқылы хабар таратты, мұның дақпырты бүкіл батыс өңірде үлкен дүмпу туғызды, тазай, қаңлы, дада-дан (тарбан), көпен секілді сегіз ел елшілерін жіберіп, бағынатындықтарын білдірді (55 б.); кіші сардар – сол кезде Да-у алайының Кіші сардары Хау Линчуан жарлық бойынша түріктерге елшілікпен барған еді... (55 б.); кіші қаған – Боғы Чордың баласы кіші қаған таққа мұрагерлік етті (55 б.); ханша – Ал олардың шын, әлде жалған ханша екенінде елдің жұмысы не? (56 б.). Бұл деректер осы күнге дейін белгісіз болып келген ғұн дәуіріне, одан кейінгі мезгілдерге жататын, бүгінгі кейбір түркі халықтарына бірқатары түсініксіз, оқылуы қиын атаулардың тізімі. Олар осы атаулар келтірілген дереккөздерден алынып, бірінші рет ғалымдар назарына ұсынылып отыр [232, 22-25].

Көнетүркі дәуіріндегі атаулар

Жалпы түркі халықтарының, соның ішінде қазақ халқының көне дәуіріндегі тарихына зерттеуші ғалымдар VI-XIII ғасырларды жатқызып, оны «көне түркі дәуірі» деп атап келеді. Біздің зерттеп отырған нысанымыз – жалпы түркі әлеміне қатысты атаулардың бәрі болмаса да, басым көпшілігі осы дәуірдегі Орхон, Енисей, Талас ескерткіштерінің тілі бойынша жасалған екі күрделі еңбекті атауға болады. Олардың біріне – ағылшындардың атақты түркітанушы ғалымы Джерар Клаусонның «XIII ғасырға дейінгі көне түркі жазба ескерткіштерінің тілі бойынша жасалған этимологиялық сөздігі» (ағылшын тілінде), екіншісіне – атақты кеңестік түркітанушылар тобының бірлесіп жариялаған еңбегі – «Көнетүркі тілінің сөздігі» (1969) жатады. Көне түркі дәуіріндегі билікке қатысты атаулар бұл еңбектерден басқа зерттеулерде толық қамтылған емес. Сондықтан да, біз қажетті атаулардың орысша аудармасы бар жоғарыда көрсетілген екінші еңбекті негізге ала отырып, көне түркі дәуірінде қолданылып, қалыптасқан атаулардың жалпы көрінісін, солардың ішінде қазақ тілінде сақталғандарын ашып көрсетуді жөн көрдік. Тарихшылардың мәліметін ескерсек, түркілердің билік тұтқаларын сипаттайтын қаған, тегін, ябғу, шад, тархан, тұтұқ, бұйрық, шор, елтебер, еркін, бек тәрізді атаулар сол дәуірдегі көне түркі тіліндегі мемлекеттік құрылым атауларының қалыптасқан жүйесін құрағаны сөзсіз [233, 13]. Біз қытай жылнамалары арқылы көне Түркі қағанаты мемлекеттілігі озық болып саналатын Қытаймен терезесі тең мемлекет болғанын білеміз. Олардың мемлекеттілігінің тағы бір белгісі – аумағының әкімшілік аймақтарға бөлінуі. Елдің ең жоғарғы билеушісі қаған атауымен белгіленді де, оң қанатқа және сол қанатқа бөлінген аймақ әкімдері – шад және жабғу деп аталды. Одан кіші тайпалық бірліктер басында елтеберлер, ал тіпті ұсақ рулар көсемдері еркін және шор лауазымдарын иеленген [234, 78]. Сонымен түркілер қоғамындағы ең жоғарғы биліктің иесі тәңір текті қаған болды. Елді бір орталыққа біріктірген қаған әрі жоғарғы билік иесі болып қана қойған жоқ, сонымен қатар жоғарғы әскербасы болып табылады. Бұл – тек тірегі мықты мемлекетке тән қасиет. Осы құқықтық сабақтастық бүгінгі Ата заңымызда да жалғасын тапты, мәселен, Елбасы және Бас қолбасшы.

Қаған сөзі – осындай мағынада қолданылған ең көнетүркілік термин. Қаған сөзінің термин ретінде қолданыла бастауы Алтай және Моңғолия аумағы секілді кең алқапта тіршілік еткен, негізінен, мал шаруашылығымен айналысатын түркі тілдес тайпалық одақтардың ерте феодалдық типті мемлекетке бірігуімен байланысты болса керек. Оның этимологиясы жайлы екіұдай пікір бар. Бірқатар түркітанушылар бұл сөзді қытай тіліне телуге өте әуес. Осы пікірді ұсынушылар қатарында кеңестік түркітанушылар А.Н.Кононов пен Н.А.Баскаков болса, еуропалық ғалымдар қатарынан Г.Дерферді, А.Габенді айтуға болады. Олардың ортақ ойынша қаған тұлғасы қытай тілінің «ке» – ұлы және «кан» – билеуші деген сөздерінен құралған. Көрнекті ғалымдардың осы пікірін қазақ тіліндегі әскери лексика құрамында қарастырған Т.Байжанов та қолдайды [235, 95]. Мысалы, түркілер қоғамында қағаннан кейінгі саяси статустары бірдей атаулар жабғу мен шад болып табылады. Көне түркілердің билеуші номенклатурасында бұлардың бірі вице-қаған қызметін атқарса, биліктің узурпацияға ұшырағанына дейін ол қарлұқтар мен оғыздарда ең жоғарғы билік титулы ретінде де қызмет етті [236, 128]. Ал екінші терминнің шығуы бұл тақтың мұрагерлігімен байланысты еді. Алғашқы тұлғаның да, соңғы тұлғаның да семантикасында орынын басушы деген мән бар, олай болса, қазіргі премьер-министрдің орынбасары, министрдің орынбасары деген мемлекеттік лауазымдарды «вице» деген кірме анықтауыш сөздермен белгілегенше, жоғарыдағы терминдердің біреуін алуға болатын секілді. Ол үшін, әрине, Премьер-министр терминіне де, министр терминіне де қазақша балама тапқан жөн. Мысалы, Ата Түріктің арқасында болған тілдік реформалар нәтижесінде қазіргі түрік тіліндегі мемлекеттік құрылыс терминологиясындағы көптеген терминдер түрікшеленді. Соның ішінде министр сөзі «бақан» деп аударылса, ал премьер-министр сөзі «бас бақан» деп қолданылады. Сонымен бірге мемлекеттік құрылыс терминологиясына қатысты деп қарайтын терминдер ішінде Н.А.Баскаков тегі қытай тілінен шыққан деп санайтын тегін сөзі бар [237, 34]. Бұл атау да түркілік антропонимдер құрамында ұшырасып, оны иеленуші адамның тегі туралы мәлімет береді. Г.Рамстедт пен Г.Дерферге сүйенген Н.А.Баскаковтың көрсетуінше, оның мағынасы текті адам деген мағына береді екен. Қазір антропоним ретінде жиі ұшырасатын еркін сөзі де жай атау болып қоймаған, ол көшпелі түркілердің саяси билік сатысында үлкен рөл атқарған. Жалпы осы сөздің құрамындағы ер сөзі түркілер санасында зор мәнге ие болғандығын С.Г.Кляшторный атап көрсетеді [238, 93]. Қазіргі қазақ тілінде де ер сөзі батыр мағынасында жиі жұмсалып жүр. Оның осындай мәні Кеңес Одағы кезінде жоғарғы награда атауының қазақшаға Социалистік Еңбек Ері деп аударылуына негіз болды. Көне түркі жазбаларында кездесетін терминнің бірі тархан – өзінің терминдік мағынасында он сегізінші ғасырға дейін жұмсалып келген ең өміршең терминнің бірі. Ол да көне мемлекеттік құрылымдардағы басқарудың бір буыны ретінде қызмет етті. Титул ретінде танылған тархан сөзінің түркілерден бастап кездесетінін Қытай жылнамашылары арқылы білеміз. Қытай тіліндегі тегі түркілік терминдердің қалайша құбылатынын осы тархан сөзінің сол тілде жазылуынан көруге болады. Бірақ қытай тілінде ол да-цзян болып өзгергенмен, А.Н.Бернштам айтқандай, «Дагян, дагань болып

өзгергенімен, оның түрік титулы тарханнан ешбір айырмашылығы жоқ» [239, 94]. Оның айтуынша, үйсіндердің өзінде-ақ ол дагян түрінде кездесіп, чиновник қызметін білдірген.

Тархан термині қазір архаизмге айналған сөз болғанымен, антропоним, этноним, топоним құрамдарында сақталып қалғанын қазақ ономастикасын зерттеген Т.Жанұзақов дұрыс көрсетеді. Мысалы, Астархан (Хаджи тархан) атының қалыптасуы туралы ғалым XIV ғасырдағы Алтын Ордаға келген араб саяхатшысы ибн-Батута, Сұлтан хажы деген түріктің қызметіне разы болып, «хаджи-тархан» деген ат беріп, қаланы алым-салықтан босатқан, содан Астархан болып кеткен деген ой айтады. Сондай-ақ ғалым қазақ тарихының материалдарына сүйене отырып, хан кезінде тархан правосын пайдаланған көптеген жерлердің осындай заңды статусына байланысты тархан атаулы топонимдердің қалыптасып кеткенін көрсетеді [240, 197]. Ол атаулардың географиялық таралымына қарап, түркілердің әр дәуірдегі мемлекеттік құрылымдарының шекарасы қай аумаққа дейін созылғанын байқауға болады.

Орта түркі дәуіріндегі атаулар

Түркі тектес халықтардың мемлекеті болып саналатын орта ғасырлардың басында құрылған мемлекеттердің бірі – Қарахан немесе қараханидтер мемлекеті. Қарахан атауын мемлекетті белгілеу үшін және осы кездегі түркі тектес тайпаларды ортақ атауға біріктіру үшін В.В.Григорьев енгізген болатын [241, 98]. Қараханидтердегі мемлекеттік лауазымдарда, негізінен, түркілік атаулар пайдаланылғаны байқалады. Мысалы, мемлекет басшысы хақан немесе тамғаш хан деп аталады, оның ортақ семантикасы «хандардың ханы» дегенге келеді. Сарай маңындағы ақсүйектер тобы «азұн тұтқушы» деген әлеуметтік атаумен берілсе, шенеуніктерінің ортақ атауы «тапұқшылар» деп аталған.

Орта ғасырлардағы кең қолданылған терминнің бірі – ұлан сөзі. Ол сол кездегі әдеби тілде оғлан тұлғасында кездеседі. Профессор Р.Сыздықтың пікірінше, түркі-моңғол тілдерінде оғлан//ұлан//ұғыл//ұл сөзінің ерте дәуірлерден келе жатқан титулдық мәнінің сарқыншағы қазақтарда Бұхар замандарына дейін жеткен тәрізді. Орта ғасырларда оғлан//ұлан сөзі – термин. Тарихшылардың айтуынша, «оғлан//ұлан деп Алтын Ордада, Қырым мен Астраханның династияларында, Чағатай ұлысы мен Ақ Ордада билеуші феодалдардың ең жоғары әскери және әкімшілік қызметін атқаратын ұлдарын атайтын болған». Соған қарағанда, ұлан сөзін әлеуметтік категорияға қатысты сөз емес, мемлекеттік ел басқару, билік жүйесіне қатысты деп қарауға болады. Қазіргі қазақ тілінде ұлан сөзі әскери-әкімшілік атау ретінде ұғынылады. Мысалы, Қазақстан Республикасының ұлттық ұланы [242, 18]. Ортағасырлық түркілік елдердің ел басқару, билік жүйесі күрделі және бай екені түсінікті. Сол себепті оларды жеке тақырыптық топтарға бөліп қарастыруға да болады. Дегенмен, біз бұл тұста жалпылама сипаттама берумен шектеліп отырмыз. Орта ғасырлардағы ел басқару және билік жүйесі біршама жоғарғы сатыға көтеріліп, күрделене бастады. Осы саладағы орта түркілік атаулар деп біздер хан, әмір, сардар, аталық, атабек, көкілташ, ішік аға басы, инақ, миршиқар, наиб, жасақ, жасауыл, қарауыл, канбұл, әкім, даруға, ұлысбек, тору, ясақ секілді сөздерді аламыз.

Ханның жарлықтары пәрмен деп аталған. Пәрмендер мәтінін тексерген зерттеушілердің айтуына қарағанда, кейде әмір деп ханнан кейінгі екінші адамды да айтқан. Әдетте, ондай адам мемлекеттің қарулы күштерінің Бас қолбасшысы болған да, ұлыс сардары қызметін атқарған. Қазіргі мемлекеттер атауына шолу жасаған кезде әмір атауын Біріккен Араб Әмірліктерінен кездестіреміз. Сонымен қатар араб елдерінің бірқатарында әмір сөзі қазір де елбасы деген мағынада қолданылады. Сонда бұл елдегі араб диалектісінде әмір сөзі әрі «әкімшілік-аумақтық бірлік», әрі «мемлекет билеушісі» деген мағынада жұмсалып тұр.

Жасауыл атауының түзілуінде де белгілі бір тілдік заңдылыққа сүйенгендік байқалады. Тілімізде осындай үлгімен жасалған қарауыл, бақауыл секілді әкімшілік мәндегі сөздер де бар. Ал енді жасауылдың одан арғы даму тарихына келсек, ол кейіннен әскери атаулар санатынан орын алып, тіпті славян тілдеріне еніп, есаул тұлғасында патшалық Ресейдің кавалериясындағы әскери шенді білдірді. Сонымен бірге орыс ұлтының құрамындағы субэтнос казактардың әскери құрамасындағы бір буын есаул атауын да түркизм деп тануға болады. Ортағасырлық ел басқару, билік жүйесіне қатысты атаулардағы тағы бір ерекшелік – көшпелі аудандардағы билік тармақтарын немесе салық түрін белгілейтін атаулар мен отырықшы жердегі билік буындары атауларының дифференциаланып қолданылуы. Мысалы, аталған елдердегі мемлекеттік істерді атқарушы хан билігінің өкілетті уәкілдері – қала әкімдері даруға деген атаумен аталады. Даруғаның міндетіне әртүрлі алым-салықтарды уақтылы жинап тұру кірген.

Жаңатүркі (Қазақ хандығы) дәуіріндегі атаулар

Қазақ хандығының қалыптасуы жайлы тарихшылар сан алуан пікірде болғанымен, біз үшін маңыздысы – сол мемлекетті сипаттайтын лексика. Алайда ұлттық мемлекетіміздің құрылуына себепші болған алғышарттар мен оған кірген ру-тайпалардың тілі туралы сөз етпейінше, жалпы сол кездегі ел басқару және билік жүйесіне қатысты атауларды толық тану мүмкін емес. Аталып отырған кезеңде түркі халықтары жеке халық болып қалыптасып, ұлттық мемлекеттер құрылды.

Профессор Р.Сыздықтың пікірінше, осы кездегі қоғамдық құрылымды көрсететін сөздерге жұрт, халық, хан, қарашы, қызметші, ер, батыр, би, ұлы, кеңес, орда секілді әлеуметтік статусты білдіретін атаулар жатады [242, 17]. Сонымен қатар, ғалымның айтуынша, әлеуметтік категорияларды көрсететін кедей, бай, патша, хан, би, төре, ақсүйек, бек, қожа, шора, азамат, батыр, ханзада, қарашы секілді атаулар жиі қолданыс тауып қана қоймай, өзінің тура мағынасында жұмсалып, сол кездегі әдеби тіл үшін актив жұмсалатын, нормаға түскен атаулар болып есептелген. Қазақ хандығындағы жоғарғы билік ханның қолында болды.

Жалпы хан сөзінің этимологиясы туралы пікірлер жоғарыда қағанға қатысты деректерге байланысты айтылған. Дегенмен, біз қаған сөзі түркі-моңғол сөзі, ал хан соның қысқарған түрі деген пікірмен бірге, көне түркі жазбаларына сүйене отырып, қаған таза түркі сөзі, ал моңғол сөзі хан да онымен бірге параллель өмір сүріп келген деген секілді ойды қолдайтынымызды ескерте кеткіміз келеді.

Тағы бір бөліп айтар мәселе – қазақ қоғамындағы хан сайлау үрдісі. Қазақ билеушілерін хан атауымен белгілегенде, оны мазмұны жағынан ерекшелеп тұратын да – осы жайт. Қазақ мемлекетіндегі хан сайлау процедурасы көшпенділер қоғамы үшін тек ұлы саяси оқиға ғана емес, сонымен қатар, үлкен этномәдени қоғамдық құбылыс болып саналған. Бұл жағдайда ел басқару және билік жүйесіне қатысты лексика этнографиялық түсініктермен астасып жатады. Мысалы, хан сайлау үшін қазақ дәстүрімен бүкіл елге сауын айтылатын болған. Бұны тек таза этнографиялық атау деп қарамасақ керек. Өйткені ол сайлау жүйесіндегі басты атрибуттардың бірі болып есептелген. Ханды таққа отырғызу кезіндегі «хан сарқыты» немесе хандықтан кетуді бейнелейтін «хан талау» тәрізді салттардың атауы ретінде де осыны айтуға болар еді. Бұлардың әрқайсысының астарында құқықтық аспект жатыр. Және осы тұрпатты жөн-жоралғылардың моральдік негізі қазіргі Конституциямызда нақты мән-мағынаға ие болып, заңды түрде көрініс тапқан.

Қазақ хандығында далалық демократияға негізделген сондай билік институтының бірі билер кеңесі деп аталған. Билер институты бұл қазақ қоғамындағы парламентаризм дәстүрлерінің құқықтық негіздерінің тым әріде жатқанын көрсетеді. Өйткені хан биліктің ең жоғарғы тұтқасы болып заң жүзінде есептелгенмен, іс жүзінде билер кеңесі оның шешімдеріне тыйым сала алатын. Лингвистикалық тұрғыдан алғанда, билер кеңесі деген күрделі атаудың құрамына қарап, сол кездегі қазақ әдеби тілінің атау-жасам қабілетінің ерте байқалғанын анықтауға негіз бар. Билер кеңесі бұған дейінгі хан кеңесі деген атаулық тіркестің үлгісінде дүниеге келген деп жорамал жасауға болатын тәрізді. Күрделі атау жасауға негіз болып тұрған модель сөз – кеңес орта түркі кезеңінен бері мемлекеттік ел басқару атауы ретінде қолданылып, кәнеш тұлғасында билеушінің канцеляриясы деген мағына берген. Қазақ хандығы тұсында үлкен ел жиындарын «кеңес» деп атаған. Айталық, «Күлтөбенің басында күнде кеңес» деген аталы сөз содан қалған. Ал би сөзінің мағынасына келсек, қазақ тілінде ол бірнеше мағына беретіні белгілі. Егер бұған дейін би сөзі әлеуметтік категорияны білдіретін бек сөзінен туындап, сол мағынада ұғынылса, алғашқы ұлттық мемлекет құрылған сәттен бастап би сөзінің мән-мағынасы кеңейді. Біріншіден, ол хан кеңесшісі деген мағынаны берді. Екіншіден, Қазақ хандығының басқа елдермен дипломатиялық қарым-қатынасының нәтижесінде елші болып баратын ресми өкілдер билерден құралатын болды. Үшіншіден, би сот билігін жүзеге асырушы адам болып есептелінді. Төртіншіден, бидің алдыңғы міндетін атқаруға қажетті қасиеттердің негізінде оның сөзге ұста, шешен деген лексикалық мағынасы қалыптасты.

Сонымен бірге Қазақ хандығы тұсында дипломатиялық лексиканың да жеке атаулар саласы болып қалыптаса бастағанын ерекше айтып өткіміз келеді. Бұл тілдік дерек те қазақтардың хандығы барлық атрибуттары бар мемлекет болғандығын дәлелдей түседі. Әсіресе, осы атаулардың ішінде елші сөзінің белсенді түрде қолданысқа енгені ерекше. «Жауластырмақ жаушыдан, елдестірмек елшіден» деген секілді көнеден келе жатқан тұрақты сөз тіркесі де еліміздің дипломатиялық дәстүрлері мықты мемлекет болғанымызды дәлелдейді. Аталған мәтелде Қазыбек бидің аузымен қалмақ ханына айтылған

әйгілі шешендік сөзінде көрінген найзаның ұшымен, қылыштың күшімен ғана емес, қажет кезінде Абылай хан секілді көрегенділікпен, сөз құдіреті арқылы бітімге келетін, жауын жұмыр жауаппен жасқай алатын, бірлікке досын бастай алатын қазақтың сыртқы саясатының негізгі стратегиялық бағыты танылып тұр. Қазақтардың, негізінен, экстенсивті мал шаруашылығымен айналысуы, көшпенділердегі экзогамиялық салттың қатаң сақталуы тәрізді жағдаяттар, сайып келгенде, олардың мемлекеттік құрылымында рулық жүйенің сақталуын аса талап етті. Өйткені рулардың әрқайсысының жаз жайлауы, қыс қыстауы ру ақсақалдарымен белгіленіп отырды, басқаша айтқанда, рулардың жер иемденуі мен олардың шекаралық межеленуінде белгілі бір жүйелілік сақталды. Рудың әкімшілік бөлініс екендігін көрсететін құқықтық белгілері болды. Мысалы, біріншіден, әр рулық ұйымның нақты аумағы белгілі болды. Екіншіден, қазақ рулары әкімшілік бөліністегі бір буын бола отырып, өзінен ірі бөлініс бірлігіне, жүздерге топтасты. Үшіншіден, рубасылар – ақсақалдар Қазақ хандығындағы беделі күшті саяси билік иелері болды. Ақсақалдар биліктің орта буынын құраған саяси институт болып қалыптасты. Бұл ХІХ ғасырдың ортасында Қоқан мен Қашғардағы болыстық басқарманың әкімшілік лауазымы болып саналды.

Қазақ хандығының отарлануы дәуіріндегі атаулар

Қалыптасқан тарихи жағдайларға байланысты Қазақ хандығының ХVІІІ ғасырдың басынан бастап, Ресей патшалығына тәуелді бола бастағаны белгілі. Осы кезде қазақтың байырғы сөздерінің мағынасы кеңейіп, кейбіреулері алғаш рет атаулық мәнде жиі жұмсала бастады. Сондай-ақ, бұрыннан келе жатқан әкімшілік пен билікке байланысты қазақтың көне атауларының орнына партия, сот, сияз, старшын, майыр, губерния, губернатор, дума, земство, окружный приказ, сенат, правление, собрание, управа, адукат секілді орыстың кірме сөздерінің қазақ тілі заңдылықтарына икемделгенімен, тұтас тұлғасында кең қолданыс табуы кең етек алды. Қазақ тіліндегі бұрыннан келе жатқан атаулардың сапалық өзгерістерге ұшырауы отаршыл патша өкіметінің қазақ даласындағы хандық билікті біржолата жоюымен байланысты болатын. Әсіресе, 1822 жылы болған «Сперанскийдің реформасы» деп аталып кеткен «Сібір қырғыздары туралы жарғы» деген құжаттың дүниеге келуіне байланысты Ресей мемлекетінің әкімшілік жүйесі ең алдымен Орта жүз аумағында, одан кейін Кіші жүзде күштеп енгізіле басталды. Сібірдің генерал-губернаторы жасаған осы жарғы бойынша бұрынғы хандық билік пен рулық жүйенің орнына аумақтық принципке негізделген сұлтандық билік орнады. Сөйтіп, қазақ даласындағы сұлтандық әлеуметтік топ атауымен бірге, жаңа отаршылдық билеу институтын белгілейтін әкімшілік атауына айнала бастады.

Қазақ қоғамындағы биліктің бір тармағы – сот жүйесі де елеулі өзгеріске түсті. Осыған орай би сөзінің мән-мағынасы өзгеріске ұшырады. Ол бұрынғыдай рудың ақсақалын көрсететін, рубасы атауы ретінде емес, жеке алғандағы қазақ даласында жаңа жағдайға байланысты туған құқығы шектеулі сот институтының атауы ретінде ұғыныла бастады. Қазақша сот билігі – билер соты шешендік сөздің күш-құдіретіне негізделген. Би – қазақ құқығы нормаларының білгірі әрі шешен болған. Шешендік өнердің бастапқы қайнар көзі,

оның сақтаушысы әрі реформаторы соттық билікті жүзеге асырған билер болатын. Осыған байланысты қазақ билері әр заманда жоғары бағаланып келеді. Бидің беделі батырдан биік саналды. «Барақ жүнді батырды екі қатынның бірі табады, Ел билейтін биді мыңның бірі табады» деген мақал осыдан қалса керек. Сонымен, қазақтың өткен замандардағы қауымында би атағына ие болған таңдаулы азамат болсын немесе қоғамдағы билік және әкімшілікке қолы жеткен азамат болсын, баршасы – жүргізген билігін қазақ құқығы негізінде, «дала демократиясының» қағидаттарына сүйеніп, логикалық парасаттылық бойынша жүргізгенін көріп отырмыз. Бұл биліктердің іс-әрекеттерінің нәтижесі олардың шешімдерінен айқын аңғарылады.

Болыс пен старшын атаулары бұрынғы хан-сұлтан қызметтеріне қарағанда ұсақ лауазым қызметтерді белгілейтін атаулар ретінде қалыптасты. Оның ең басты себебі – патша өкіметінің жергілікті биліктің құзырын шектеп, ұсақтату арқылы далалық метрополиядағы өзінің саяси билігін күшейту пиғылынан болатын. Сол тұрғыдан келгенде, атаулардың анықтамаларының өзінен-ақ патша өкіметінің отаршылдық саясаты айқын білінеді. Соның нәтижесінде Қазақ хандығы көп жағдайда өзінің қалыптасқан құрылымдық мәртебесін, әдеттегі атақ-лауазымын, хан мемлекетінің құрылымын, билік жүйесін толық сақтай алмады. Кей кезеңдерде хандық лауазымы аға сұлтанмен, облыс басшылары – волостной управитель, генерал-губернатор және т.б. атаулармен алмастырылды;

Қазақ тіліндегі билікке қатысты атаулар халқымыздың тарихи-этнографиялық үдерістері, рухани мәдениеті, әдет-ғұрыптары мен салт-дәстүрлері, дүниені қабылдауы және т.б. жөніндегі нақты ұғым-түсініктерді береді. Оларда тіл мен этностың, тіл мен рухани мәдениеттің өзара тығыз байланысы жарқын көрініс тапқан. Билікке қатысты сөздер – жекелеген тілдерге ғана тән тіл бірліктері емес, этникалық жағынан туыстас, материалдық және рухани мәдениеті ортақ бірнеше халықтың тілінде қолданылатын этномәдени атаулар. Халқымыздың қоғамдық, саяси-әлеуметтік өмірінде қалыптасып, оның рухани, мәдени, азаматтық тарихында ғасырлар бойы өз қолданысын тауып келе жатқан бұл атаулар және оларға қатысты жүздеген, тіпті мыңдаған тұрақты тіркестер мен мақал-мәтелдер тіл байлығымыздың сүбелі саласын құрайды. Бұл тіл байлығын лингвомәдени ішкорпусқа енгізу – бүгінгі таңдағы өзекті мәселелердің бірі болып табылады. Әзірлену үстіндегі лингвомәдени ішкорпус тілдік бірліктер арқылы қазақ тілінің қолданысын кеңейтетін цифрлық платформаның бір үлгісі, тіл білімі, әдебиеттану, әдістеме және т.б. салалардағы зерттеулерге қажетті мәтіндер базасы бола алады. Сонымен қатар бұл лингвомәдени ішкорпус қазіргі кезде көптеген тілдердегі түрлі стильдегі мәтіндерді қамтитын (<https://opus.nlpl.eu/>) және т.б. платформаларды, яғни лингвомәдени ішкорпустарды жасаудың бастамы деп санаймыз. Келтірілген тілдік бірліктер қазақ тіліндегі мәдени-танымдық ақпараты бар этномәдени бірліктердің мәдени семантикаларын ашу жұмыстарына, энциклопедиялық сипаттағы репрезентативті лингвомәдени ақпарат беру, этнобірліктердің лексикографиялық базасын әзірлеуге қажетті материалдар жинау, мәдени-танымдық мазмұндағы мәтіндер базасын әзірлеу жұмыстарына қосылатын үлес деп санаймыз.

8.9 Диаспора қазақтары тіліндегі этномәдени бірліктер

Кез келген тілде болсын, не ұлттық тілдің басқа формалары – диалекті, сөйленісте болсын ондағы түрлі сыртқы және ішкі лингвистикалық себептерге байланысты орын алатын тілдік өзгерістер лексикалық қорынан айқын көрінеді. Олай дейтініміз, дыбыс жүйесінде, морфологиясында, синтаксисінде, т.б. құрылымдарда өзгерістер болғанымен, олар өте баяу жүреді. Ал лексикалық қорда әрбір 15-20 жыл сайын жаңа сөздердің пайда болып, қайта жаңғыруы, қолданылу жиілігі артуы немесе сөздердің ескіріп, салғырт қорға өтуі, қолданыстан шығуы сияқты үдерістердің қарқындылығы жоғары. Сондықтан қазақ тілінің сөйленістеріндегі, шетелдегі қазақтардың тіліндегі жергілікті, аймақтық лексикалық бірліктер үнемі жинақталып, арнайы диалектологиялық сөздіктерде көрініс табуы тиіс. Аталмыш сөздер халқымыздың әр өңірлердегі тұрмыс-тіршілігі, салт-дәстүрі, кәсібі, шаруашылығы, өмірлік тәжірибесі, өзге халықтармен қарым-қатынасы мен қолданылу сипатынан мағлұмат беретін тіліміздің байлығы болып табылады. Бұл салада Академик Ә. Қайдаров [243], Р. Сыздық [244], Е. Жанпейісов [245], Н. Уәлиев [246], Ж.Манкеева [247], Р. Шойбеков [248] т.б. отандық ғалымдар үлкен үлес қосты.

Қытайда тұратын қазақтардың сөздік қорында қазақ елінде ұмыт болған немесе басқа мағынада қолданылатын қазақ халқының заттық және рухани мәдениетін айшықтайтын сөздер мен сөз тіркестері сақталған. Атап айтқанда, мал ауруларына байланысты: *көз, жылқы обасы, уақұрт, қанқұрт, сіреспе, таңдайқақ, бақана, алаөкпе, солма, сый, қысаға, тіл тікен, жемір, ақ болу, кегежсек, сарамас, көтеу, бөрітас, сапақ т.б.* Салт-дәстүрге байланысты: *жыртыс той, жыртыс сый, жыртыс буу, жыртыс апару, жыртыс жырты, балдыз көрімдік, босағаға ілу, қыз қашар, қыз көтерер, келіннің атына міну, есік көре келу, күйеу аттандырар, бақан салды, қапқа салар, күйеу аяқ, мойын тастар, қолтық сүйер, ит ырылдатар, қыз құшақтар, күйеу аттандыру, қыз шығару т.б.* Ет жасап, қонақ күтіп, табақ тарту салты бойынша: *бас табақ, қос табақ, келін табақ, беташар табағы, жерік асының табағы, сый табақ, ошақ табақ, ерулік табақ, сарқыт табақ, қыз алар(табақ), орта табақ, аяқ табақ т.б.* көптеген кәде, салт-дәстүрлер осы уақытқа дейін өзге елде өмір сүріп жатқан қазақтардың тыныс-тіршілігінде сақталған, қолданыста жүрген дүниелер. Әсіресе, Қытай және Моңғолия сияқты қазақтар шоғыры отырған өңірдегі қазақтардың тілдік қолданысы мен өмір салтында бар осындай мәдени құндылықтарымызды «жергілікті ерекшелік» деуге келмейді, көбісі сонау заманнан келе жатқан дәстүрлер саналады. Сондықтан да оларды жалпы қазақ халқының заттық және рухани мәдениетін айшықтайтын сөздер мен сөз тіркестері, салт-дәстүрлері қатарына жатқызуға болады. Біле білсек, олардың біразы Қазақстанның да кей өңірлерінде сақталған. Мәселен, Қазақ хандығы тұсындағы басқару институттары саналатын *тағанақ кеңес* пен *тұрымтай кеңес* сөздері Илияс Есенберлиннің, ал, *тұяқ таға, бұлыншақ, көзілдірік* сөздері Ақселеу Сейдімбеков шығармаларында кездеседі. Бүгінгідей ұлт мәдениетін ұлықтау, «рухани жаңғыру» үдерісі жүріп жатқан тұста шетелдегі қазақтардың өзін ғана әкелмей, сонда қалып бара жатқан сөзін де әкеле

алсақ, нұр үстіне нұр болатыны даусыз. Тіпті, кейбір дәстүрлер «қара жорға биі» сияқты қайта жаңғыруы әбден мүмкін. Жиналған мұндай сөздер мен сөз тіркестері болашақта зерттеулерге пайдасын тигізіп, сөздіктердің мазмұнын байыта түсуге өз үлесін қосары да сөзсіз.

Қытайда тұратын қазақтардың сөздік қорында этномәдени бірліктермен бірге кірме сөздер, баспасөздерінде ресми қолданыста жүрген аймақтық бірліктер мол кездеседі. Өзге жұрттармен арадағы ұзақ дәуірлік қарым-қатынас нәтижесінде солардың тілінен енген, бұл күнде дыбыстық жағынан қазақ тілінің ыңғайымен түбегейлі өзгеріске ұшыраған мәселен, қытай, моңғол, маньчжур, тибет, санскрит, араб, парсы, ұйғыр, орыс т.б сөздері де молынан кездеседі. Бұл сөздердің бірқатары – қазақтардың сөйлеу тілінің дыбыс жүйесіне бейімделіп, дыбыстық өзгеріске түскен (қыт. сәй – көкөніс, борми – жүгері, лазы – бұрыш, шедеу – көкөніс турауға арналған үлкен пышақ т.б.). Біразы айтылуы мен жазылуы жағынан түбір негіз нұсқасын сақтап, кірме сөзге айналған. Кірме сөздердің қосымша морфемалары да, қазақ тілінің дыбыс жүйесіне бейімделіп, өзгеріске ұшыраған. Жергілікті тіл ерекшеліктері ретінде қарастырылған тілдік бірліктер өңір қазақтарының тілінен, баспасөз, интернет көздерінен, көркем әдебиеттерден, жазба деректерден жинап жүйеленді. Ондағы тілдік қолданыста тарихи сөздер (*тәйжі, амбы, бейсі, үкірдай, залың, зәңгі, гуң, дамолла т.б.*), кәсіби сөздер (*жемір, жаққұрт, бөрітас, кегежек, көз, қанқұрт, қаракемік, қортық, қын, сипай, солма т.б.*), жергілікті терминком бекітіп, баспасөздерінде ресми қолданыста жүрген аймақтық бірліктер (*бейзаттық мәдениет, бейүкіметтік ұйымдар, бейөндірістік тарау, жаңалық жарату технологиясы, туындайтын энергия, білімділік экономикасы, қолфон, Еуропа ортақ тұлғасы, мемлекеттік істер жорасы, бас шужи, үкіметсіздік, аманат-қарыз, күміс лоңқа, базар экономикасы, куәші, қамсыздандыру, қамсыздандыру серіктестігі, серіктестік жасау, мемлекеттік жолдама, дипломатиялық тектемелеу, уәкілдер үйірмесі, орынбасар зуңли, ақша базары, қазына құрылымы, ерен базар, төрешілдік, аткезшілік, аткезшілік тауарлар, шипагер, шипахана, келтехат, көмек ақша, мұзсандық, өспе, торап, торап дос, тіл тақта, тор бекет, торап жарнама, сұхбат үні, тораптық қазына, тораптық талдағыш, торап сөздігі, торап ескертпесі, тінтуір(мышка), жұмсақ бөлшек, қатты бөлшек, тіркегіш, десте, бетбелгі, бекет бастығы, жершарылану, бақылап-тізгіндеу, сөздер кестесі, тәжірибе атызы, артқы есіккен жүру т.б.) лексикалық ерекшеліктердің болуы қазақтардың рулық құрамы, орналасу аумағы сонымен бірге қытай тілі мен мәдениетінің, қоғамдық-саяси жүйесінің ықпалы екендігі анықталды.*

Қытайда тұратын қазақтардың тілінде отандық сөздіктерге әлі ене қоймаған тілдік бірліктер молынан кездеседі. Кейбір сөздер мен сөз тіркестері түсіндірме сөздіктерде кездескенімен, онда үлкен айырмашылықтар бар. Мәселен «маяуза» сөзіне он бес томдық «Қазақ әдеби тілінің сөздігінде»: «*Киізден жасалған жертөсеніш немесе жеке адамның астына төсеуге арналған дөңгелек, шаршы пішіндегі шағын киіз төсеніш*» деп берілсе, Қытайда тұратын белгілі жазушы Шәмсі Құмарұлы: «*Маяуза – серкенің күзгі қалың терісін илеп, жасы үлкен адамдар үшін жасалған төсеніш. Кейде туған ту ешкінің терісінен де маяуза жасайды*» деп түсіндірген [249].

Сол сияқты «кәзекей» сөзіне сөздікте: *«Жеңі жоқ, жағасы кеңірек ойылған, белден төменірек түсіп тұратын, мақпалдан тігілген әйелдер киімі»* деген түсініктеме берген. Ал біз көрген материалдарда кәзекейді: *«Кәзекей – жеңі жоқ, жағасы жоқ, мықыннан төменірек түсіп тұратын жеңіл киім, ең бастысы, оны тек әйелдер ғана емес ерлер, балалар да киеді»* деген. *«Балалар кәзекейі – қыстық, жаздық болып бөлінеді. Жаздық кәзекейлер түрлі кездемелерден жалаң астарлы, жанқалтасыз, омырауынан үш түймелік орнатып, қаусыратын етіп тігіледі. Қыстық кәзекейлер – тері кәзекей, сырма кәзекей»*, – деп олардың әрқайсысына жеке анықтамалар берген.

Сондай-ақ 15 томдық сөздікте «қоя» сөзіне: *«Кұстың бөтегесі мен иттің асқазанында қортылмай, жеген асынан қалып қойған жарамсыз зат (жүн-жұртқа, қоқсық т.б.)»* деген анықтама берілген. Біз оқыған материалдарда «қоя» тек құс пен иттің ғана асқазанында ғана емес, кәдімгі малдың да асқазанында болатындығын айтады. Сондықтан кітапта «қоя» деген сөзге *«малдың жеген жем-шөбінің асқазанында қорытылмай, түйдектеліп қалған қалдығы (топырақ, құм түйіршіктері, жүн-жұртқа, қыл-қыбыр т.б.)*. Қоя аң-құстардың асқазанында да болады. Қоясы бар мал жақсы тойынбайды, кейде іш аурудың белгілері көрінеді, малдың сыртынан қояны анықтау қиын, тек сойғанда ғана білуге болады» деп толығырақ анықтама бердік.

Сол сияқты «қанжілік» деген сөзге 15 томдық сөздікте *«бәйгеге қосуға жаратылмаған, майға бөккен, семіз»* деген анықтама берілген. Нұрлан Қинаятұлы *«Қазақтың дәстүрлі мал шаруашылығы мен мал шипагерлігі»* деген еңбегінде оны *«жылқы, сиыр, түйе сияқты малдардың жілік майының бұзылуынан жілігіне қан түсіп, жілік майына араласуынан болатын ауру. Жылқыларды жас кезінен демалдырмай мінуден, бәйгеге көп қосудан, ал түйе мен өгіз болса, ауыр жүк артудан, қатты шаршап, зорығудан қанжілік болады»* деп түсіндіреді [5].

Кітапқа енгізілген сөздердің басым бөлігі Қазақстанда жарық көрген сөздіктерде кездеспейтін, Қытайда тұратын қазақтардың тілдік қорында сақталған немесе сол жерде жарық көрген қазақ басылымдарында көрініс тапқан сөздер мен салт-дәстүрлер. Атап айтқанда, мал ауруларына байланысты: *көз, жылқы обасы, уақұрт, қанқұрт, сіреспе, таңдайқақ, бақана, алаөкпе, солма, сый, қысаға, тіл тікен, жемір, ақ болу, кегежсек, сарамас, көтеу, бөрітас, сапақ т.б.* Салт-дәстүрге байланысты: *жыртыс той, жыртыс сый, жыртыс буу, жыртыс апару, жыртыс жырту, балдыз көрімдік, босағаға ілу, қызқашар, қызкөтерер, келіннің атына міну, есік көре келу, күйеу аттандырар, бақан салды, қапқасалар, күйеу аяқ, мойынтастар, қолтық сүйер, итырылдаатар, қыз құшақтар, қыз шығару т.б.* Ет жасап, қонақ күтіп, табақ тарту салты бойынша: *бас табақ, қос табақ, келін табақ, беташар табағы, жерік асының табағы, сый табақ, ошақ табақ, ерулік табақ, сарқыт табақ, қыз алар(табақ), орта табақ, аяқ табақ т.б.* көптеген кәде, салт-дәстүрлер осы уақытқа дейін өзге елде өмір сүріп жатқан қазақтардың тыныс-тіршілігінде сақталған, қолданыста жүрген дүниелер. Мысалы:

ДӘМ АЙҚАСТЫРУ э т н. *Қазақтың неке салтында қыз жақтың жігіт жаққа сөз салып, құда түсу дәстүрі. Қыз жақтың әке-шешесі біреудің ұлын*

жақсы көріп, қызын соған бергісі келсе, жігіттің үйіне арнап дастарқан алып, қызын ерте барады. Дастарқанға төстік, асық жілік сияқты кәделі жіліктер түйеді, «дәм айқастырғалы келдік» деп жасырмай айтады. Жігіт және жігіттің әке-шешесі қызды жақтырса, дәмін алып қалады, жақтырмаса, дастарқанын шешпей сол күйі қайтарады. Қазақ жолында мұны «д ә м а й қ а с т ы р у» дейді (Ясын Құмарұлы «Қазақтың салт-дәстүрі»).

ДӘМГЕ ТАПСЫРУ э т н. «дәм ұрсын!» деген мағынада. «Балапан басына, тұрымтай тұсына» деп, ол үшін саған тырнақтың көлеміндей бүкпем болса, д ә м г е т а п с ы р (Қ. Мұқажанұлы. Тайталас).

БЕТАШАР ТАБАҒЫ э т н. Қазақтың келін түсіру салтында, беташар салты орындалғаннан кейін, келіннің бетін ашқан жігітке арнап тартылатын табақ. Беташар табаққа тіл жағын, жүрек, бауырын, ортан жілік, жауырын, белдеме, арқа, қабырғаларын, үстіне таңдай салады. Оның үстіне әдемі кестелі қол жаулық жабады, болса күміс жүзік, теңге қояды. Бет ашқан жігіт жолдас-жораларымен алқа-қотан отырып, теңгені, жүзікті, қол жаулықты бөлісіп алады да, бата жасайды, әнші жіпт таңдайым тақылдасын деп таңдайды алып жейді, қалған еттерді бәрі бірігіп бөлісіп жейді. Мұны қазақ кәдесінде «б е т а ш а р т а б а қ» деп атайды (Болат Бопайұлы «Беташар табақ»).

ЖАМБАС СИПАР э т н. о й ы н. Қыз-бозбалалардың айлы кеште сан санап, белгілі санға келгенде жаңылпаш айтып, жамбас сипап ойнайтын ойын түрі. Ұшына орамалдың түйдім сусар, Көп жылқы көк алалы көлде жусар. Кешегі ел қыдырған есер шақта, Ән еді Біржан салған «Ж а м б а с с и п а р» (Біржан сал).

ЖЫРТЫС АПАРУ э т н. Күйеу жігіт жақтың тойдан бұрын қыз жаққа келін киетін киім-кешек, көрпе-жастық, түскііз сияқты заттарды апару салты. Кей жерлерде «жасау апару», «жүк апару» деп те атайды. Ж ы р т ы с т ы а п а р ғ а н кезде, қыз ауылы шашу шашып, қуана қарсы алады (Ясын Құмарұлы «Қазақтың салт-дәстүрлері»).

ЖЫРТЫС БУУ э т н. Жыртыс той өтіп, қыз ауылына жыртыс апарарда той иесінің шағын тілеу жасап, туыс-туған, дос-жарандарын шақырып, құда жаққа апаратын сыйлықтарды қоржын-қолаң, теңдерге салатын дәстүрі. Ж ы р т ы с б у ғ а н д а, көбінде, әрбір теңге лайығында торғын-торқа, дәм және басқа да ұсақ заттар қоса буылады (Ясын Құмарұлы «Қазақтың салт-дәстүрлері»).

ЖЫРТЫС ЖЫРТУ э т н. Жігіт жақтан болашақ келінге келген жасауды шешу кәдесі. Жігіт жақтан келген құдағидың бірі қыздың шешесіне бас теңді көрсетіп, одан әр теңде қандай жыртыс бар екенін айтып, алып келген бұйымдарын түгел тапсырып береді. Ал, қыздың шешесі туыс-туған, дос-жарандар мен жегжат-жұрағаттардың әйелдерін шақырып, тең шешіп, жыртыс жыртады. Бас теңді ауылдағы сыйлы, жолы үлкен әжелер шешеді, оған салынған ж ы р т ы с т ы ж ы р т у ғ а келген әйелдер жолына, жасына қарай отырып бөліп алады [250].

КЕГЕЖЕК з а т в е т. Қозы, лақтарда болатын ұстамалы ауру, «қозы талмасы» деп те атайды. Қозы к е г е ж е к болғанда денесі сіресіп, басы кекжиіп, көзі алайып, аузынан көбік ағады [251].

БҰРҚЫЗАТТАР. *Ел ақтаған дуана. Шанишылды от басына асатаяқ, Үкі таққан әлемдеп бастан аяқ. Ала сеңсең бөркі бар, тысы аққиіз, Есіркейді ел бұрқыны көрген аяп* (Т. Жолдыұлы).

БҰЛЫНШАҚЗАТЭТН. *Ыстық күнде айран-шалап, қымыз сияқты сусындықтарды құюға арналған, жылқы терісінен жасалатын ыдыс. Бұлқыншақты жасау үшін жылқы сойғанда артқы санының терісін шашасына дейін мес қылып бітеу сойып алады. Содан соң терінің көк етімен шелін тазартып, сан жағын арасына шүпірек салып, тарамыспен немесе жылқының құйрық қылымен тігеді. Тігіс бұлқыншақтың түп жағына келеді. Терісінің сирақ тұсына келетін тар жағына лайық ағаштан тығын жасалады да ішіне күл толтырып, тығындап көлеңкеге кептіреді. Әбден кепкен тері қатып көнге айналады. Содан соң күлін төгіп тастап, ыдысты ыстайды, ары қарай ыдысты шаймен жуып, ішіне бірнеше күн ашыған айран құйып қояды, оны «шикі дәмін алу» дейді. Одан кейін пештің құрымын жинап алып, оны суға малып алып, теріні біраз уақыт соған салып қояды. Оны «құрымдау» деп атайды. Содан соң ыдыстың ішіне қайнаған сұйық май құйып, шайқап алып әрі қарай пайдалана беруге болады. Бұлқыншақты керегенің басына бауынан іліп қоюға болады. Салт атпен жүргенде ердің алдыңғы жағына ілінеді немесе қанжығаға байланады. Жаяу адам иығына асып жүреді. Бұл ыдыс малышы қауымның ыстық күндерде айран, шалап, қойыртпақ құюына мейлінше пайдалы. Бұлқыншақты жасау үшін жылқы сойғанда, артқы санын шашасына дейін мес қылып, бітеу сойып алады.*

КӘКИМӘЕТІКЗАТЭТН. *Бойжеткен қыздар мен жас келікшектерге арналып, былғарыдан оюлап, сәндеп тігілген биік өкшелі етік. Ою-өрнек салған көксауыр етік пен кәкимә етік бір киер ретінде той-томалаққа киіледі (Әбдіжәлел Оразбайұлы «Қазақтың дәстүрлі киім-кешектері»). Жарғақ шалбар, түйе жүн шекпен, кең құлаш етік, ақ тандақ, Қызыл шибарқыт қамзол, аяқтарында қызғылт кәкимәетік, ақ патсайы шытырман көйлек (Ілияс Есенберлин «Қаһар»).*

КЕЛІННІҢ АТЫНА МІНУЭТН. *Ұзатылған қыз алғаш келін болып ауылға келгенде орындалатын салт. Ұзатылып келген қыз ауылға таяғанда, аттан түсіріліп басына желек жабылады да, оның мініп келген атына жолы жақсы, өмірі баянды, балалы-шағалы, ибалы әйелдердің бірі мініп, елден бұрын ауылға жетеді. Оны ауыл адамдары шашу шашып қарсы алады. Келіннің атын мінетін әйелге әлгіндей талап қойылатын себебі, келінде соның ізін бассын, ұрпақты болсын, бақытын тапсын, дәулеті шалқысын деген ырымнан туған. Келіннің атына мінген адам көрпенің арасына салынған «тақым тартар» деген кәдені алады (Ясын Құмарұлы «Қазақтың салт-дәстүрлері»).*

Әсіресе, Қытай және Монғолия сияқты қазақтар шоғырлы отырған өңірдегі қазақтардың тілдік қолданысы мен өмір салтында бар осындай мәдени құндылықтарымызды «жергілікті ерекшелік» деуге келмейді, көбісі сонау заманнан келе жатқан дәстүрлер саналады. Сондықтан да оларды жалпы қазақ халқының заттық және рухани мәдениетін айшықтайтын сөздер мен сөз тіркестері, салт-дәстүрлері қатарына жатқызуға болады. Біле білсек, олардың бірсыпырасы Қазақстанның да кей өңірлерінде сақталған. Мәселен, Қазақ хандығы

тұсындағы басқару институттары саналатын *тағанақ кеңес* пен *тұрымтай кеңес* сөздері Илияс Есенберлиннің, ал, *тұяқ таға*, *бұлыншақ*, *көзілдірік* сөздері Акселеу Сейдімбеков шығармаларында кездеседі. Бүгінгідей ұлт мәдениетін ұлықтау, «рухани жаңғыру» үдерісі жүріп жатқан тұста шетелдегі қазақтардың өзін ғана әкелмей, сонда қалып бара жатқан сөзін де әкеле алсақ, мұндай сөздер мен сөз тіркестері болашақта зерттеулерге пайдасын тигізіп, қазақ тілінің корпусы мен сөздіктердің мазмұнын байыта түсуге өз үлесін қосары да сөзсіз.

8.10 «Түйе» сөзінің семантикалық тіркесімділігі және ұлттық корпуста автоматтандыру тәсілдері

Ойды қалыптастыруға негізгі механизм болатын белгілер – менталитет, этникалық мәдени бейне, дүниенің этникалық бейнесі болмақ. Бұл белгілі бір этникалық топтың ұлттық тілін, ана тілін сақтауға қызмет етеді. Демек, тілдік санада дүниенің этникалық бейнесі сақталған азаматтардың ұлтқа тұрақтылығы, ұлттұтастырушылық құндылығы жоғары болады. Ғаламның этникалық бейнесін, этникалық мәдени бейнені вербалдайтын тіл бірліктерінің әлеуетін зерттеу, лексикографиялау – қазіргі қазақ тіл білімінің өзекті мәселесі. Этномәдени ұғымдар мен атаулардың мәдени кодын анықтау үшін жаңа талдау тәсілдері қажет. Соның бірі – семантикалық талдау. Біздің ұсынғалы отырған әдісіміз – семантикалық валенттілік әлеуеті арқылы тілтаным жүйесінде сөздің ұғымға айналуы, сөздің мәңгілік ұлт жадында сақталуына оның тіркесуінен туындаған ұғымдардың вербалдық қызметі басымдыққа ие болады екен.

Зерттеудегі екінші мәселе, ұлттың дүниетанымында этникалық ұғым ретінде сақталатын атауларға талдау үлгілері ұсынылады. Мәселен, қазақ – мал шаруашылығымен күн көретін халық. Соның ішінде ең бағалы мал – түйе. Ол шөлге төзімді, еті мол, сүті пайдалы, жүні киім-кешекке, үй құрылысына жарамды, пайдасы мол жануар. Қазақтың климатына бейімделген жануар болғандықтан, қазақ халқының негізгі тіршілік көзіне айналған. Сондықтан қазақ тілінде 100-ден аса түйеге қатысты тіркестер бар. Бұл – қазақ мәдениетінде этникалық ұғым немесе лингвокультурема ретінде қабылданады. Лингвокультуремадан құралған этнографизмдерді цифрлық ресурста, ұлттық корпуста қалай әзірлеу керек? Іздеу жүйесінде «түйе» атауына қатысты фитонимдер, зоонимдер, антропонимдер, этнографизмдер мен мифоморфизмдер, фразеологизмдер мен метафоралар, коннотаттық мағыналарды жеке ұяшықтарда рәсімделуі арнайы компьютерлік лингвистика әдістерімен жүзеге асады.

Зерттеуімізде үш мәселеге қатысты ұғымдардың өзара байланысы арқылы тілтанымдық өнімнің жүзеге асу үдерісі баяндалады. Сондықтан, біріншіден, семантикалық тіркесімділік мәселесіне тоқталайық. Тіл біліміндегі тіркесімділік сөздің басқа сөзбен тіркесуінде өзіне жақын семантикалық үйлесімділігін бірінші кезекке қояды. Бұл жайында Г.И.Кустова былай дейді: «Тілдік жүйедегі семантикалық ығысуда сөздің бастапқы мағынасына айтарлықтай тәуелділік болады. Себебі одан тарайтын семалық мәндер мен реңктер туындайды. Оларды туынды мағынаға жатқызады, негізінде олар бастапқы мағынаның семантикалық топтары немесе басқа тіркесімділікке ие болған

атаулар деп саналады. Семантикалық тіркесімділік мағына ауысуының басты көрсеткіші» [252, 74]. Тілдегі тіркесімділікті сол тілдің тұтынушысы ғана сәтті тіркестіре алады. Бұл жайында И.Г. Казачук пен Л.П.Юздова былай дейді: «Тіркес сөз жасаушы ең белсенді процессуалдық бірліктер – етістіктер мен фразеологизмдер. Етістіктің тіркесімділігі актанттардың белгілі бір шеңберімен біріктіру қабілетін білдіреді. Тіркесімділік әртүрлі тілдік деңгейлермен байланысты: семантикалық (сөз мағынасының құрылымы және фразеологизм), морфологиялық (көмекші), синтаксистік (предикат). Сондықтан тіл валенттілігі ұғымын жүйелі түрде зерделеу қажет» [232, 126].

Ю.Д. Апресянның зерттеуінде: «Семантикалық тіркесімділік сөздің лексикалық мағынасымен тікелей байланысты және оны басқалардан ерекшеленетін нақты лексикалық бірлік ретінде сипаттайды. С. Fillmore терминдерін қолданғандағы (субъект, объект, құрал, орын және т. б.), осындай нақты түсініктер лексикалық мағынаның бір мәні [254, 56]. Бельгиялық тілтанушы К. Виллемс етістіктің тіркесін тіл мен логика тұрғысынан қарастырады. Оның пікірінше: «Тілді әлем туралы адамның аялық білімі тұрғысынан, яғни жалпы семантикалық таным тұрғысынан қарастырған жөн. Себебі тіркесімділік семантикалық рөлдер мен оның күрделі құрылымдары арасындағы аралық деңгейде орналасқан. Мағына үйлесімділігін таңдау валенттілік құрылымдары мен оның перифериялық бөлшектеріндегі мазмұндық сәйкестікке тікелей байланысты. Ал лексика тіркес құрылымдарын семантикалық деңгейдегі ақпаратпен байланыстырады» [255, 56]. Демек, сөздердің семантикалық тіркесуі адамның бейнелі ойлау, заттарды салыстыру, ұқсату қиялынан, коннотатты ойлау қабілетінен туындайды. Яғни тілдегі этникалық дүние бейнесі сөздердің валенттілігі арқылы вербалданады. Байқағанымыздай, семантикалық тіркесімділік адам жадында терең сақталатын тұрақты тіркестердің, фразеологизмдер мен бейнелі тіркестердің негізін қалайтын ұғымның ғана тіркесім перифериясын құрай алады. Демек, семантикалық тіркесімділік әлеуеті жоғары тірек сөз – дүниенің этникалық бейнесін таныта алатын сөз деп саналады. Сондықтан дүниенің этникалық бейнесі ұғымына да тоқталайық.

Тіл білімінде дүниенің этникалық бейнесі біріншіден, ұлттың менталды сипатын (характерін) айқындайды, екіншіден, ұлт ретіндегі қалыптасу, тұтасу феномені ретінде қызмет етеді, үшіншіден, өзге этностық топтардан ерекшеленетін мәдени, психологиялық, рухани, менталды, философиялық және көркемдік танымдағы айрықша белгілерін көрсетеді.

Келесі кезекте, ұлттық корпусның базасын рәсімдеу тәсілдеріне қатысты лингвистикалық зерттеулерге шолу жасаймыз.

1. Бірінші орында – Орыс тілінің ұлттық корпусы (<https://ruscorpora.ru/>) [5]. Себебі, онда 2 миллиард сөздік қоры бар. 16 ішкорпусы бар (негізгі корпус, ауызша корпус, газет корпусы, синтаксистік корпус, әлеуметтік желі корпусы, акцентологиялық корпус, мультимедиялық корпус, мультипарк корпусы, параллель корпус, диалектілер корпус, поэтикалық корпус, оқу корпусы, орыс классиктерінің корпусы, тарихи корпус, панхрониялық корпус, от 2 до 15 корпусы).

2. Екінші қатарда 100 миллион сөздің мәтіндік базасы бар Британдық ұлттық корпус артықшылығы – лексикографиялық базасы кең, екіншіден,

жазбаша корпус -90%, жан-жақты метабелгіленімдері бар, бірақ ішкорпусы аз (<http://www.natcorp.ox.ac.uk/>) [256].

Қазіргі таңда барлық ұлттық тілдердің корпусық жүйесі іске қосылып жатыр. Себебі бұл – қоғам қажеттілігі. Сондықтан корпусық лингвистика қарқынды дамып келе жатыр. В.П. Захаровтың пайымдауынша: «Кейбір ғалымдар бүгінгі таңда корпусық лингвистиканың тіл білімі саласындағы негізгі бағыттардың бірі деп таныса, кейбірі корпусық лингвистиканы тіл білімінің дербес саласы ретінде қарастырады» [257]. Ұлттық корпус – тілдердің тарихи тілдік қорын сақтауға, тіпті оны қайта тірілтуге негіз болады. Мәселен, Ресей Федерациясындағы башқұрт, қалмық, якут тілдерінің тұтынушылары азайып барады. Оларды сақтау үшін жеке ұлттық корпусы құрылды. Біздің зерттеуімізше, ұлттық корпус – ұлттың этникалық тіл бейнесін танытатын электронды ресурс. Сондықтан ұлттық корпус – лексикографиялық қорды сақтау жүйесі. Сондықтан әрбір мағыналық талдау ұлттық корпус базасында семантикалық белгіленімдермен көрініп отыруы тиіс. Осы мақсатта «түйе» сөзіне қатысты қазақтың этникалық тілдік бейнесін ұлттық корпуста рәсімдеу жолдарын көрсетеміз.

Дүниенің этникалық бейнесі тілтанымдық ІТ ресурстарда көрсету үшін мынадай әдістер қолданылды:

1. Әлеуметтік-философиялық зерттеу әдісі. Белгілі бір ұлттың әлеуметтік тіршілігіндегі ұлтқа тән философиялық пайымның ұлттық құндылыққа айналуын, құндылықтың ұрпақтан-ұрпаққа жалғасып, жойылмауы, ұлттың сана жадында сақталуын зерттеу. Мәселен, қазақтың этномәдени танымында түйе – кеңдіктің, қуаныштың, молшылықтың, төзімділіктің символы. «Түйенің жарты етіндей» теңеудің мағынасы – үлкен денелі, алып адам. Осы тұрақты тіркестің орнына үлкеннің жалпыадамзаттық символы болатын – тау, биік ағаш неге алынбады? Бұнда қазақ халқының таным-түсінігі, үлкендікті қабылдау стереотипі бар.

2. Мәдениеттанымдық зерттеу әдісі. Ю.Караулов ұсынған «семантикалық гештальт» [258, 25] жүйесі бойынша дәстүрлі ортада қолданылатын материалдық және рухани мәдениеттің ырым-тыйымдарда, жосын-жоралғыларда қолданыс жиілігінің арқасында тілдік жүйеге сіңісуі, яғни мақал-мәтелден, фразеологиялық оралымдардан орын табуы анықталады. Мәселен, қазақ халқы «үлкен қуаныш», армандап жүріп жеткен бақытты «ақ түйенің қарны жарылды» деген фразеологизммен жеткізеді. Қазақ мәдениетінде ақ түйе – әбден бапталған, тойдың асына дайындалатын түйе. Сондықтан ақ түйені сойып, етінен түрлі тағам жасап, той өткізу – ең үлкен қуаныш деп танылған.

3. Жиілік талдау әдісі. Қазақтың сөздік қоры мен лексикографиялық базасында түйеге қатысты сөз тіркестерін, күрделі сөздерді санадық. Түйе сөзінен жасалған 588 тіркес бар екені анықталды. Бұл жиілікті анықтауда А.Байтұрсынұлы атындағы Тіл білімі институтының картотекалық қоры (<http://tbikartoteka.kz/>) [259] сайтының іздеу жүйесіне сүйендік.

4. Тақырыптық модельдеу атты компьютерлік лингвистика әдісі қолданылды. Түйе сөзіне қатысты бірнеше тақырыптық ұяшықтар ашылды. Семантикалық және валенттілік белгіленімдерді рәсімдеу жолдары анықталды.

5. Сөздің тіркесімділігін анықтау әдісі – сөзжасамның семантикалық және лексикалық тәсілінің бірі. «Түйе» сөзімен тіркесуге әлеуеті бар етістіктер мен зат есімдер таңдалды. Олардың тіркесуінен туындаған күрделі сөздер мен сөз тіркестерінің предикаттық, номинативті, субстантивті қызметтері ажыратылды.

Байқағанымыздай, жоғарыдағы 5 әдіс «Түйе» сөзінің тіркесімділігін анықтай отырып, одан тараған жаңа сөздердің этномәдени бірлік ретінде лексикографиялық базаға тіркелуін және ұлттық корпуста жеке сөздің леммасы бойынша іздеу парақшасын әзірлеудің жолдары ұсынылды.

Кез келген тілдің лингвомәдени атаулары мен этнографизм сөздерінің тіркесуі әлеуеті жоғары болады. Себебі ол сөздерде мәдени код бар. Әсіресе, этномәдени атаулардың көне символдық, эталондық, ритуалдық, мифтік мағыналарын танытатын семантикалық өрісі кең болады және ол сол тілдің лексикографиялық құндылығы болып терең сақталады. Бұл зерттеуімізде жеке ұғым, жеке сөз ретінде сөздік құрауға лайықты бола білген сөздерді ұлттық корпуста әзірлеуге болатынын дәлелдегіміз келеді.

Біріншіден, «Түйе» сөзіне қатысты концептуалдық топтарды анықтадық:

1. Түйе – Орта Азиядағы халықтық және культтық жануар. Себебі түйе шөл, шөлейтті климатта өмір сүреді. Бұндай аумақта Орта Азия халықтары орналасқан. Сондықтан осы өңірде өмір сүретін барлық түркі тілдес халықтардың тіршілігінде түйеге қатысты культ бар. Сондықтан туристерге арналған түйе кәдесыйы – Азияның эмблемасы деп танылады.

2. Түйе – қазақ мәдениетінде үлкендіктің, іріліктің символы. Қазақ халқында үй жануарлары – төртеу. Түйе, жылқы, сиыр, қой. Соның ішінде ең ірісі – түйе. Басқа жануарлармен салыстырғанда да үлкендігінен түйе көзге түседі. Себебі түйе – шөпқоректі динозаврлардың сарқыншағы болып саналады да *Aerycamelus*, *Paracamelus* [260, 47] ұрпағы деп танылады. Сондықтан қазақ тілінде «түйе кесе» «түйекұс» атаулары үлкендікті білдіру мақсатында түйе тіркесімен беріледі.

3. Түйе – ұлттық стереотиптердің тілдік көрінісіндегі лингвомәдени атау.

Біріншіден қазақ мәдениетінде түйе – көлік қызметін атқарушы жануар. Алыс шақырымдағы ағайын-туысқа түйемен жеткен. Мысалы, «түйе шана». Сондай-ақ түйе – жүк тасымалдаушы көлік. Мысалы: «шаңырақ түйе» –көшпелі қазақтың киіз үйін артқан ауырлыққа төзімді түйе.

4. Түйе – дүниенің этникалық бейнесін танытатын этномәдени атау.

Қазақ мәдениетінде «түйе» сөзіне қатысты ою (орнамент) атауы, ыдыс-аяқ атауы, ауылшаруашылық атауы, ұлттық ойын атауы, халықтық өлшем атауы мен фитонимдер, зоонимдер мен поэтонимдер кездеседі. Мысалы: түйетабан – бұл ою атауы. Қазақ халқы жамбы, төсеніш, кілемге осы оюды ырымдап салатын болған. Себебі түйе тоқшылықтың, молшылықтың нышаны деп түсінілген. Сондай-ақ «түйе көш жер» өлшем атауы – 50 шақырым қашықтықты білдіреді. Демек, түйе ауыр жүкпен 50 км жерде демалмастан жүк тасымалдай алады. «Түйеқуырдақ» – тек түйенің ғана емес, кез келген еттің үлкен етіп туралып, майға қуырылған ұлттық тағамы. Бұнда түйе еті – негізгі акцент емес, түйенің үлкендігіне ұқсатылған ет бөлігінің үлкендігі прецедент болған.

Екіншіден, ұлттық корпусстың ішіндегі «ТҮЙЕ» сөзін рәсімдеудің құрылымдық жүйесін ұсынамыз:

1. «Түйе» сөзінің анықтамалық белгіленімі. Бұл ұяшыққа түйе сөзінің энциклопедиялық анықтамасы мен бейнесі енгізіледі.

2. «Түйе» сөзінің лексикалық белгіленімі. Онда бірнеше метабелгіленім болады:

2.1. Лексикалық мағына. ТҮЙЕ – бір немесе екі өркешті, күйіс қайтаратын ірі мал, үй жануары.

2.2. Зоологиялық мағынасы. ТҮЙЕ – сүтқоректілер табына сірітабандылар отрядына жататын күйіс қайтаратын жұптұяқты бір немесе қосөркешті ірі жануар.

2.3. Культтік мағынасы. ТҮЙЕ – еті мен сүті азыққа, терісі мен жүні әртүрлі тұрмыс қажетіндегі бұйымдарды жасауға, өзі шөлге төзімді болған соң мініске пайдаланылатын, иесінің дәулеттілігінің белгісі ретіндегі үй жануары.

2.4. Көне мағынасы. ТҮЙЕ – ерте кезде Шығыс елдерінде жүк тасуға, ұзақ жолға, шөлге төзімді болғандықтан, сауда керуендерінің көлігі.

2.5. Этнографиялық мағынасы. ТҮЙЕ – қалыңмалға берілетін тоғыз сый-құрметтің басты белгісі.

2.6. Діни мағынасы. ТҮЙЕ – Ислам діні бойынша құрбандыққа шалуға болатын, ең ірі мал.

3. «Түйе» сөзінің тіркесімділік белгіленімі. Бұл ұяшықтан «түйе» сөзімен тіркесетін 4 түрлі қызметтік топ ашылады:

3.1. Атау тіркесімділігі. «Түйе + зат есім» тіркесімі арқылы жасалған атау тіркестер енеді. Яғни түйе сөзімен тіркесетін зат есімді атаулар. Әрбір номинативті тіркестің мағынасы беріледі.

Түйе шелек – сұйықтық құйылатын үлкен ыдыс. Түйе ассоциациясы – үлкен мағынасын білдіру үшін дүние бейнесіне айналған.

Түйе қымыз – күзді күнгі түйенің сүтінен ашытылып дайындалатын су-сын.

Түйе бас – дене қимылына арналған ойын, бір адамның үстіне бірнеше адам мініп ойнау. Кім үстіне көп адамды мінгізіп, көтере алса, сол жеңеді. Семантикалық басымдық түйенің ауыр жүк көтере алатын қасиетінде болып тұр.

Түйе жүн көрпе – жылуды сақтайтын, түйе жүнінен жасалған көрпе. Семантикалық доминанты – түйе жүнінің физикалық, химиялық қасиетіне байланысты, яғни жылусақтағыштық қасиетіне қарай, қыста қолданылатын қалың жамылғы.

Түйе сүйек – сеңгіртаспа, бұршақ тұқымдасы, таспа туысына жататын бұта. Бұл сөз тіркесі де халықтың ассоциациясы бойынша жасалған.

Түйе бүлдірген – бүлдіргеннің үлкендігі бармақтай, тұтас алша сияқты бір түрі. Мағыналық ассоциациясы – үлкендік, ірілік.

3.2. Предикативті тіркесімділік. Ішкорпустың бұл ұяшығында «түйе» сөзімен тіркесетін етістікті атаулар іріктеліп беріледі. «Түйе + етістік» тіркестері ұсынылады да мағыналары ашылады.

Түйе шөгерді – түйе тізесін бүгіп, жатты.

Түйе қамдады – түйеге жүк арту үшін арқасына киіз салды.

Түйе бас қылды – бәрі жиналып басып, жаншып тастады.

Түйеге қом бітті – түйе семіріп, салмақ қосты.

3.3. Анықтауыштық тіркесімділік. Ішкорпустың бұл ұяшығында «түйе» сөзімен тіркесетін атрибутивті (анықтауыштық) қатынастағы күрделі атаулар іріктеліп беріледі. «Анықтауыш сөз + түйе» тіркестері ұсынылады да мағыналары ашылады.

Бая түйе (ашуланшақ, бақырауық түйе) – қасына тірі жан жақындағанды ұнатпайтын, ыңырсып, қарсылық білдіретін түйелердің бірі.

Шау түйе – жүк көтеруден қалған, денесі кішірейе бастаған әлсіз, кәрі түйе.

Атан түйе – піштірілген түйе.

Жабы түйе – өмірі мінілмеген, қашаған, асау түйе.

Ауған түйе – бөтен жерден мекенін ауыстырған түйе.

Алмас түйе – екі өркеші екі жағына жығынқы түйе.

Жарық түйе – көп мінілген, тіркеуден мұрны жарылған түйе.

Жегін түйе – шаруашылыққа, арба жегуге, мініске, жүк артуға үйретілген түйе.

3.4. Фразеотіркесімділік. Ішкорпустың бұл ұяшығында «түйе» сөзімен тіркесетін фразеологизмдер беріледі және мағыналары ұсынылады.

Түйенің құмалағындай – ірі, үлкен.

Түйені түгімен жұтты – параны көп алу, тойымсыз.

Түйе үстінен ит қапты – икемсіз, епсіз, болбыр.

Түйені мұзға жетелеу – бос әуреге түсу, болмайтын іспен айналысу.

Түйенің танытыны – жапырақ – білетіні бір-ақ нәрсе.

Түйенің тұяғы түсетін жер – өте алыс жер.

4. «Түйе» сөзінің тақырыптық-семантикалық белгіленімі. Ішкорпустың бұл ұяшығына түйе сөзі қатысқан барлық тіл бірліктері (588 бірлік) семантикалық-тақырыптық топтарға іріктеледі.

4.1. Фитонимдер. «Түйе» сөзінен басталатын өсімдік атаулары тізбектеледі.

Түйебеде – түйенің сүйсініп жейтін шөбі.

Түйекөтен – қызанақ. Бұл сарказммен айтылған диалект, халықтық атау. Қызанақтың ең ірісі, етті помидорды қазақ халқында осылай атаған екен.

Түйесіңір – жазғытұрым мал жейтін өсімдік. Сабағы ұзын бұта болғандықтан, түйенің ірілігімен салыстырған қазақ халқы осылай атап кеткен.

Түйесораң – қаңбақ. Майда шөптерді жел домалатып, шар тәрізді жиынға айналдырады. Сол қаңбақ шөптің сыртқы пішіні түйенің мұрнынан ағып жерге түскен сораңға немесе аузынан ағып жерге түсіп домаланып жатқан түкірікке ұқсағандықтан түйесораң деп атаған.

Түйесүйек – бұтағы қатты, желге төзімді, шайқалмайтын қақайған өсімдік. Майыспайтын қатты жуан бұтаны түйенің сүйегіне ұқсатудан туындаған.

Түйетікен – сабағы ұзын болып өсетін тікенді қурай. Бұл өсімдікті түйе сүйсініп жеп, күйіс қайтарады. Соған орай қазақ халқы осылай атаған.

Түйетабан – түйенің табанына ұқсас тарбиып өсетін шөп

Түйеқарын – сабағы бүдір-бүдір, біресе жіңішке, біресе жуан болып келетін өсімдік. Сабағының ұзындығын, сабағының жуан жері түйенің қарнындай жуан, томпақ болуын ұқсату негізінде аталған.

Түйекұйрық – тікенекті биік қарасора. Түйенің сүйсініп жейтін шөбі
Түйекекіре – тікенекті кекіре. Түйенің сүйсініп жейтін шөбі
Төмендегі өсімдік атаулары үлкендігіне, ірілігіне қарай түйеге ұқсатылып теңестіруден қалыптасқан:

Түйешетен – үлкен жидек

Түйебұршақ – үлкен бұршақ

Түйежантақ – үлкен жантақ

Түйежаңғақ – үлкен жаңғақ

Түйеошаған – үлкен ошаған

Түйежапырақ – жапырағы өте үлкен өсімдік

Түйешырмауық – үлкен шырмауық

Түйебүлдірген – үлкен бүлдірген

Байқағанымыздай, түйе сөзі фитонимдер, біріншіден, түйенің сүйіп жейтін өсімдіктерінің атауларында, екіншіден, түйенің денесінің үлкендігіне, ұзындығына, қарнының үлкендігіне, түкірігінің үлкендігіне ұқсатып, яғни түйенің сыртқы пішіні мен бейнесіне ұқсатып қойылған өсімдік атауларында жиі кездеседі екен.

4.2. Зоонимдер. «Түйе» сөзінен басталатын жануар атаулары тізбектеледі.

Түйетек – Лама. Қазақта «лама» – түйетек деп аударылады. Себебі, қазақтың дүниетанымында алғаш ламаны көргенде түйе тұқымдас жануар іспетті стереотипті қалыптастырса керек.

Түйеқұс – Страус. Қазақ мәдениетінде страус – үлкен құс ретінде қабылданған. Ал қазақта үлкендік түйемен стереотиптенгенін байқаймыз.

Түйекене – Ең үлкен кене.

Түйекиік – Гуанако. Сыртқы пішіні түйеге ұқсайтындықтан осылай атаған.

Түйетабан – Балық. Түйенің табанына ұқсаған жалпақ балық.

Түйеқоңыз – Үлкен қара қоңыз. Малдың жас тезегін домалатып, зыңылдап ұшып жүретін, басы түйенің басына ұқсаған қоңыз.

Түйеқұлақ – Тышқан. Құлағы түйенің басына ұқсаған үлкен тышқан.

Түйе шыбын – Денесі ірі, үлкенірек шыбын.

Түйешағала – мойны ұзын, денесі ірі түйе секілді сарғыш қоңыр түсті шағала.

Түйекекілік – түйе керуені секілді, бірінің сонынан бірі жұптасып жүретін құс.

Қорытындылай келе, байқағанымыз, жануар атауларында түйе сөзінің жиі кездесуі ең алдымен түйенің сыртқы пішіні (үлкендігі), қоңыр сары түсі, мойнының ұзындығы, басының үлкендігі, сондай-ақ түйенің керуен қызметі де ассоциацияланатынымен түсіндіріледі.

4.3 Оронимдер. «Түйе» сөзінен басталатын географиялық ландшафт атаулары тізбектеледі.

Түйе жон – түйенің арқасындай еңкіш келген төбе дала.

Түйетомар – үлкен, биік төмпешік жер.

Түйеқорым – ірі тастардың құлауынан пайда болған қорым.

Түйемойнақ – түйенің мойны сияқты иіліп келген таудың аласа жері.

Түйетартар – түйе секілді сіміретін, су соратын оқпан.

Түйешаңлақ – шөл дала.

Қорытатын болсақ, тау, жер, төбе, жазық дала атауларында да түйе сөзінің кездесуі, түйенің шөл, шөлейтті зонада мекендеуі, су ішкенде шөлін қандырғанша көп әрі ұзақ сіміріп соратыны, түйенің мойны мен өркешінің тау-төбені сипаттауы қазақ халқында түйенің сырт пішіні ғана емес, өзіне тән кимылдары, өзгелерден айрықша белгілері терең ассоциацияланып, дүниенің этникалық бейнесіне айналған.

1.3. Тұрмыстық бұйым атаулары. «Түйе» сөзінен басталатын бұйым атаулары (ыдыс, жиһаз, құрал, т.б.) тізбектеледі.

Түйебел – тоқуға арналған жуан ши.

Түйе кесе – үлкен кесе.

Түйе шығыр – түйеге жегіліп айналдыратын шығыр.

Түйе қазан – түйе еті түгел сыйып кететін үлкен қазан.

Түйемойнақ – дорба. Түйенің мойын терісінен жасалған су құятын дорба-ыдыс.

Түйе мойнақ орақ – орақ. Сабы түйе терісінен жасалған орақ.

Түйе өркеш қалып – қалып ыдыс.

Қорытатын болсақ, тұрмыс бұйымдарында түйе сөзінің кездесуі түйенің мойын терісінің төзімділігі, су өткізбейтін қасиетімен танылған, содан ыдыс жасау негізінде бірқатар ыдыс атауларында түйе сөзі кездескен. Сондай-ақ үлкен қазан түйе еті түгел пісіретін қазанмен емес, түйенің бейнесімен ассоциацияланған. Бұнымен қоса түйенің өркеші сопақ қалыпқа ұқсатылып, ыдыс, бұйым атауында түйе сөзі стереотиптелген. Егер біз түйе сөзінен басталатын сөздерді жинап, осылайша семантикалық-тақырыптық топқа іріктесек, қазақ халқы үшін түйе сөзінің тілдік құндылығын, дүниенің этникалық бейнесі ретінде қабылданатынын байқамас едік. Тіл біліміндегі дүние бейнесі мәселесі талданған мақалада: «Тілдегі ұлттық-мәдени элемент арнайы басқалардан өзгеше формалды көрсеткішке ие болмайды, бірақ сол тілдік бірліктің мағынасында астарлы (имплицитті) ішкі семантикалық форманы қалыптастыра алады» [261, 780], – делінеді. Демек, дүниенің этникалық бейнесін құрайтын этнографизмді анықтау үшін ішкорпус ұяшықтарына жинақтап, ұлттық корпуста рәсімдеу аса маңызды. Себебі дүниежүзіндегі қазақ диаспорасы үшін қазақтың ұлттық мәдениетін терең сақтау қажет және әлемдік мәдениетте де қазақтың этникалық танымы туралы ақпарат таралуы тиіс. Қазақстандық ғалым А.Баймурзинаның пікірінше, «Қазақстан көп ұлтты ел ретінде толерантты қарым-қатынас үшін жағдай жасауға ұмтылады» [262, 801].

«Түйе» макротоп ұяшықтарын түйеге қатысты ас-тағам атауларымен, ою-өрнек (орнамент) атауларымен, тоқыма атауларымен, киім-кешек атауларымен, опа-далап атауларымен, сусын атауларымен толықтыруға болады. Бұл мақаланың көлемі көтермейтін болғандықтан қалған ішкорпус ұяшықтарын ашып көрсете алмадық. Дегенмен, оларды болашақта (<https://qazcorpus.kz/>) [40] сайтынан көруге болады.

Байқағанымыздай, дүниенің этникалық бейнесі – жалпыадамзаттық дүние бейнесін толықтыратын құрамдас бөлігі. Этникалық бейнені құрайтын ақпараттар – этностың өзіне ғана тән, өзге мәдениетте жоқ культтер. Әсіресе

өзге мәдениетте жоқ жосын-жоралғылар қазақтың ұлттық тілдік бейнесін танытуға қызмет етеді. Этникалық дүние бейнесі адам санасында тұрақталған сайын ұлттық тілдің бейнесі кеңейеді. Мәселен, қыздың бұрымды болуы – сұлулық эталоны десе, жалпыадамзатқа тән әлемдік тілдік бейне болады. Ал «қыз болсын деп, соңғы туған ұлға тұлым қою, керісінше, ұлға зәру болғандар соңғы туған қызға айдар қойып, ұлша киіндіріп өсіруі» қазақтың этникалық бейнесін танытады. Осындай ешбір мәдениетте жоқ, этникалық таным-түсініктерді тілдік жүйеден іздеу тың зерттеуді қажет етеді. Бір ұлттың этникалық бейнесін ашу олардың әрбір танымдық дүниесінің мәдени кодын ашумен тығыз байланысты.

Талқылау барысында «Түйе» тіркесімділігінен құралған күрделі атаулар дүниенің этникалық бейнесін танытушы концептілікті өрісті құрай алды. Оны былайша көрсетуге болады:

Қазақтың этникалық дүние бейнесін танытатын «ТҮЙЕ» этномаркері.

Түйе – іріліктің маркері: Түйе балтыр. Қазақ мәдениетінде адамның балтырын суреттеуде, адамның сыртқы келбетіне портрет жасауда ірі денелі адамды сипаттауда міндетті түрде түйе бейнеленеді. Мәселен, бойы ұзын, ірі денелі адамның балтырын суреттеуде түйенің аяғымен теңестіре ассоциациялау бар.

Түйе – төзімділік маркері: Ақ иық түйе – жүк көтеруден шаршамайтын, еңбекке төзімді адамды сипаттайтын фразеологизм. Ешкімнің көңілін қалдырмайтын, жұмсағаннан бас тартпайтын, шаршағанын білмей жүк тасытын адал жұмыскерлерді осылай түйенің қасиетіне ұқсата отырып суреттейді.

Түйе – көшпелілер салт-дәстүрінің маркері: Жігіт түйе – күйеу жігіт үйленбей тұрып қалыңдықтың атасына сыйлық береді. Оны қазақ мәдениетінде «ұрын бару» дейді. Жасырын барғанда қыздың әкесіне семіз түйені арнайы баптап сыйлыққа ала барады. Сондай салт-дәстүрге арнайы бапталатын түйені «жігіт түйе» деп атаған. Демек, түйе – үй жануары ғана емес, сыйлық ретінде де ұғынылады.

Байқағанымыздай, түйенің семантикалық тіркесімділігі арқылы түйе туралы халықтың концептілік аялық білімін тануға мүмкіндік алдық. Тірек сөздің атрибутивті, предикаттық, номинативті тіркес сөздері болмаса, нысанға алынған ұғымның дүние бейнесін аша алмас едік. Демек, тірек сөзге тіркесетін сөздерді де халық таңдап тіркестіреді. Сол тіркес сөзді таңдауда ұғымның мазмұнын ашатын, ішкі бейнесін дәл бере алатын сөздер іріктеледі. Мысалы, қоң торғай түйе. Күтімі жоқ, азыққа жарымаған, арық, кішкентай түйе деген мағынаны білдіреді. Яғни еті жоқ, құр сүйегі қалған жүдеу түйені торғайдың етіне теңеуі литота әдісімен бейнеленген. Сөз тіркесінде де семантикалық тіркесімділіктің беріктігі жоғары болғаны соншалық фразеологизмге айналған. Демек, семантикалық тіркесімділігі сәтті үйлескен сөз тіркестері константалық деңгейге жетіп, паремиологиялық қордан орын алады.

Қорыта келгенде, қазақтың этникалық бейнесін тілдік жүйеден анықтау үшін:

– біріншіден, жалпыадамзаттық концептілік ұғымдарға элеуметтік-философиялық зерттеу жүргізу қажет. Сол арқылы материалдық ұғымдарды

сипаттауда халықтың танымын, халықтың ассоциациясын, халықтың аялық білімін, халықтың ұқсату заңдылығын анықтауға мүмкіндік туады. Бұл дүниенің этникалық бейнесін жасайтын концептілік құрылымдарды айқындап береді.

– екіншіден, ұғымның материалдық немесе рухани мәдениетке немесе құндылыққа айналғанын тану және зерттеу маңызды. Себебі семантикалық тіркесім астарында, яғни сөздердің тіркесуінен туындаған күрделі атаудың мәні ұғымның концептілік әлемін танытуға қызмет етеді.

– үшіншіден, семантикалық тіркесімділікті мағыналардың үйлесімдігі тұрғысынан қарастыру маңызды. Себебі семантикалық тіркесімділігі сәтті болған күрделі атаулар сөздік қорда терең тұрақталады. Бірі фразеологизмге айналады, бірі теңеу немесе эпитет болып сақталады. Егер тіркескенде сөздердің мағыналары магнит секілді бірін-бірі тартпаса, онда бұл сөз тіркесін тіл тұтынушылары қолданбайды немесе қолданыстан шығып ұмытылады. Семантикалық тіркестердің гармониясы дегеніміз – сөздердің мағыналық сәйкестігі әрі фонетикалық үйлесімділігі, морфологиялық үйдесімділігі. Жасанды аффистермен жасалған екі сөздің тіркесі сәтті бола алмайды. Егер екі сөзді тіркес болса, алдыңғы сыңардың бір-екі буынды болуы сәтті болады, ал үш сөзді тіркес болса, бірінші сыңары бір буынды, екі-үш әріпті болуы сәтті болады.

– төртіншіден, семантикалық тіркесімділікте символдық құндылығы бар сөздердің қатысуы маңызды. Сол арқылы ұлттың немесе этностың ең қадірлі, қастерлі, киелі, құрметті ақиқат дүниелері, оның бейнесі анықталады. Символдық мәні бар сөздің тіркесімділігінің қолданысы жоғары болады.

– бесіншіден, қазақтың этникалық дүние бейнесін анықтау үшін, мифологиялық, стереотиптік, зооморфтық, фитонимдік, прецеденттік этномаркерленген атаулар мен материалдық мәдениеттің санада этномаркерлену құбылысын зерттеу этномәдени атаулардың мәдени кодын ашуға ықпал етеді.

9 МАҚАЛ-МӘТЕЛДЕР ІШКОРПУСЫ

9.1 Мақал-мәтелдер ішкорпусы: мәтіндік базасы, түсіндірме беру әдістері

Корпус – бүгінгі таңда ұлттық сипатта көлемі өте үлкен мәтіндердің автоматты түрде, компьютерлік бағдарламалар арқылы жұмыс істейтін, жалпы көпшілік пен тілтанушы ғалымдардың ғылыми зерттеу жұмыстары мен оқытушы-мұғалімдерге, білім алушыларға оңай және тез ақпарат алуға мүмкіндік беретін таптырмас инновациялық-ақпараттық құрал ретінде компьютерлік тілдік қорды құрайды [263].

Компьютерлік лингвистика мамандарының тұжырымдауынша, компьютерлік тілдік қор дегеніміз – ғылым адамының өз зерттеу нысанына жаңаша тұрғыда көз салу мүмкіндігі [264], [265]. Мұндай тілдік қор неғұрлым қомақты (көлемді) болса, солғұрлым тіл құрылысының сыры тереңірек ашылады, зерттелетін нысан жөніндегі түсініктердің аумағы кеңейеді. Сол сияқты, зерттеуші адамның мүмкіндік-қабілеті әлденеше есе артады, шығармашылық қуат көздері ашыла түседі, сөйтіп, бұл жаңа мүмкіндіктер қазақ тілінің жүйесін жетілдіруге және тіл жүйесін мұқият тануға жұмсалатыны сөзсіз. Сонымен бірге тілдік процестерді автоматтандыру мәселесі тоқырап қалмай, үнемі жетілдіру жағынан өзінің жалғасын тауып тұруы қажет.

Бүгінгі жаһандану заманында ғылым-білімнің барлық саласын автоматтандыру кезек күттірмейтін өзекті мәселеге айналып отыр. Тілдік материалдарды автоматтандыру қажеттігі қазіргі кезде тіл білімінің «Компьютерлік лингвистика» атты саласының дамып, қалыптасуына мүмкіндік берді. Тіл білімінің жаңа саласы болып табылатын «Корпустық лингвистика» – «Компьютерлік лингвистиканың» бір саласы.

Корпустық лингвистика компьютерлік технологияларды қолдана отырып, лингвистикалық корпустар (мәтіндер корпустарын) құрастыру мен оны пайдаланудың жалпы ұстанымдарын зерттейді.

Корпустық лингвистика саласының зерттеу пәні – әртүрлі салалардағы пайдаланушылардың тілдік зерттеулер жүргізуіне арналған тілдік деректердің көлемді ауқымын жасау мен оны пайдаланудың теориялық негіздері және практикалық механизмдері [266].

Ал лингвистикалық немесе тілдік, мәтіндер корпусы дегеніміз – нақты тілдік мәселелердің шешімін табуға арналған аса үлкен көлемдегі мәшине (компьютер) оқи алатындай түрде көрініс табатын, бірыңғайланған, құрылымдалған, белгіленген (шартты белгілер қойылған), филологиялық тұрғыда компетентті саналатын тілдік деректер ауқымы. Басқаша айтқанда, қазақ тілінің мәтіндер корпусы (Қазақ тілінің ұлттық корпусы) – аса ірі көлемдегі жүйелі түрде құрылымдалған, тиісті лингвистикалық және экстралингвистикалық (метамәтіндік) белгіленімдер енгізілген нақты тілдік мәселелерді компьютер арқылы шешуге арналған тілдік деректер ауқымы (массиві) [267].

Мәтін таңдау, корпуста мәтін енгізу үдерісінде жинақталған тілдік материалдарға қатысты репрезентаттық (репрезентативность) пен теңгерімділік

(сбалансированность) мәселелері туындайды. Әсіресе бұл мәселе Ұлттық корпустарды қалыптастыру кезінде аса қажет. Корпустың репрезентативтігі – мәтіндік материалдардың көлемі мен әртүрлілігі жағынан ықтималдықтар теориясының қанағаттандыру шартына бағынады.

Осы ретте қазақ әдеби тіліндегі әртүрлі функционалдық стильдерді (көркем проза, драматургия, ғылыми-техникалық мәтіндер, публицистикалық мәтіндер) қамтитын лингвистикалық ақпаратпен толық аннотацияланған (мазмұндалған) ұлттық корпусты құрастыру – бүгінгі таңда аса өзекті мәселе.

Корпустар өз ішінде субкорпустарға бөлінеді: поэзия тілінің корпусы, проза тілінің корпустары, газет мәтіндерінің корпусы, ғылыми-техникалық мәтіндер корпусы, ұлттық корпус т.б.

Корпус авторларының міндеті – корпус мақсатына сәйкес салалардан барынша көп мәтіндер жинау. Корпусты сол тілдің кішірейтілген моделі деуге болады. Репрезентативтік ұғымы – барынша жеткілікті және әртүрлі кезеңдердегі мәтіндердің жанрларға, стильдерге, авторларға және т.б. қатысты пропорционал шамамен алынған көлемі, яғни тілдік зерттеу аясын жан-жақты ашатын мәтін көлемі [268].

А.Н.Барановтың тұжырымдауынша, репрезентативтікті жүзеге асырудың ең қарапайым жолы – пәндік салалар бойынша алынатын таңдама мәтіндерді сәйкес деңгейде шағындау әдісі. Міне, тек осы жағдайда ғана корпусты ұйымдастыруда сәйкестік стратегиясы сақталатын болады [264].

Сонымен, репрезентативтілік (жан-жақтылық) – корпустың пәндік аяның барлық қасиетін бейнелей алу мүмкіндігі деуге болады яғни лингвистикалық зерттеу типіне қатысты пәндік аядағы құбылыстың кездесу жиілігі тілдік бірліктерді бір-бірінен ажырата алатындай мәнде болуы қажет [265].

Корпустық лингвистика ең кем дегенде екі түрлі типтегі нысандармен (мәтіндер корпустарымен) әрекет етеді:

1. Бірінші типтегі корпустар әмбебаптық сипатта болады, олар тілдік әрекеттің барлық саласын қамтиды.

2. Екінші типтегі корпустар арнайы мақсатпен жасалған, мысалы, мақалдар корпусы немесе газет тіліндегі саяси метафоралар корпусы [265].

Екі жағдайда да репрезентативтік критерийі сақталады.

Бірінші түрге жататын мәтіндер корпустары 60-жылдары әмбебап сипатта жасалды. Мысалы, АҚШ-та сол кезеңге сай қанағаттанарлық репрезентативтік дәрежеде Браундық мәтіндер корпусы құрастырылған болатын. Оған 15 жанрды бейнелейтін 6-дан 80-ге дейінгі қарапайым таңдама мәтіндер іріктеліп алынады: 1) баспасөз; 2) пресса: бас мақала; 3) пресса: шолу; 4) діни мәтіндер; 5) дағды, сабақ, әуестік (хобби); 6) ғылыми-көпшілікке арналған әдебиет; 7) беллетристика, өмірбаяндар, эссе; 8) әртүрлі (үкіметтік құжаттар, кәсіпорын есептері, өнеркәсіп есептері, колледждердің каталогтары); 9) ғылыми шығармалар; 10) көркем әдебиет; 11) мистика мен детективтер; 12) ғылыми проза; 13) қызық оқиғалы әдебиет және вестерндер; 14) сүйіспеншілік жайлы роман-дар; 15) әзіл-оспақ шығармалар [269].

Екінші типтегі корпустарда репрезентативтілік критерийі ретінде нысанға алынған құбылыстың қолданыста барын барынша объективті түрде

бейнелеу мақсаты қойылады. Мәселен, ағылшын тілін иеленушінің сөйлеу тәжірибесіндегі белгілі бір кезеңдегі және географиялық аймақтағы максимал репрезентативтікке ие ағылшын тіліндегі мақалдардың корпусы [270], [271].

Қазақ тіліндің стильдері бойынша ғана емес, ұлт тілінің сөз байлығы болып табылатын басқа да тілдік бірліктерінің – тіркестердің, фразеологизмдердің, мақал-мәтелдердің корпусын жасау аса қажеттілік тудырады.

Мақал-мәтелдер корпусының дереккөзі қазақ тілінде жинақталған, кітаптарда жарық-көрген материалдардан алынған. Өйткені мақал-мәтелдер – тілімізде көне дәуірден қалыптасып, дайын тұрақты тіркестер ретінде қалыптасып, қорланған тілдік бірліктер болып табылады. Сондықтан оқылымда бар, бірақ корпус қазақ тілінде мақал-мәтелдер корпусы алғаш рет жасалып отыр.

Қазіргі кезде мақал-мәтелдер корпусында 2000 мақал-мәтел уәждік түсіндірмесімен және ағылшын тіліндегі балама аудармасымен енгізілген.

Корпустық менеджер – бұл жүйе корпустағы деректерді іздеп-табуға арналған бағдарламалық құралдарды өз қатарына қосуды бірге, статистикалық ақпаратты алуды және пайдаланушыға алынған нәтижелерді ыңғайлы түрде ұсынуды көздейтін мамандандырылған іздестіру жүйесі болып табылады. Корпус бойынша зерттеуші өзіне қажетті сөздерді, сөз тіркестерін, грамматикалық формаларын арнайы құрастырылған іздестіру жүйесі арқылы іздеп-тауып, олардың мағыналарын аша алады.

Қазіргі кездегі корпустық менеджерлердің көпшілігі ақпараттың бірнеше түрін іздеуді жүзеге асырады: нақты сөзформаларды іздеу; леммаға сәйкес келетін сөзформаны іздеу (бір ғана лемманың мәтінде кездескен барлық формаларын іздеу); тұрақты және тұрақсыз сөзтіркестерді іздеу.

Мақал іздеу ұяшығына кіру арқылы пайдаланушы мақал-мәтелдер корпусынан тірек сөз арқылы қажетті мақалдар легін таба алады.

Мақал-мәтелдер корпусы мақалдар құрамындағы сөз және сөзформадан, тақырыптық топтан іздеу жүйесі бойынша жұмыс істейді. Қазақ тіліндегі тақырыптық, лексика-семантикалық топтарға қатысты зерттеулер біршама жасалған. Мақал-мәтелдерді әртүрлі мақсатта пайдаланушы адам корпустан өзіне қажетті сөз қамтылған не тақырыптық топқа қатысты мақал-мәтелдерді қазақша уәждемесі мен ағылшынша аудармаларымен бірге іздеу ұяшықтарына жазу арқылы тез тауып алуына болады.

Мақал-мәтелдер корпусының дереккөзі ретінде қазақ тіл білімінде мақал-мәтелдерді этнолингвистикалық тұрғыдан зерттеген академик Әбдуәли Қайдар еңбектері негізге алынды. Ғалымның мақал-мәтелдердің этнолингвистикалық семантикалық мағынасын ашқан әдісі әдіснамалық негіз ретінде басшылыққа алынып, әрбір мақалдың шығуына себеп болған ұғымдар мен жағдайлар уәждемелік түсіндірме ретінде беріледі [272].

Мақал-мәтелдер корпусы тек мақал-мәтелдер тізімі емес, мақал-мәтелдердің түсіндірмесі берілген және ағылшын тіліне аударылған іздеу жүйесімен жұмыс істейтін ақпараттық құрал. Сондықтан біз мақал-мәтелдерге этнолингвистикалық талдау жасаған Ә.Қайдар еңбектерін дереккөз ретінде алып отырымыз. Сондай-ақ «Бабалар сөзінде» қазақ тіліндегі мақал-мәтелдер көлемі

жағынан көп жинақталғандықтан, «Бабалар сөзінен» жинақталған мақал-мәтелдер осы этнолингвистикалық әдіспен түсіндіріледі [220].

Қазақ тілінде мақал-мәтелдердің түсіндірме уәждері бұрын Ә.Қайдардың зерттеулеріне дейін арнайы жасалмаған болатын. Қазіргі кезде мақал-мәтелдерді сирек қолдану үрдіс алып бара жатқаны өкінішті жайт. Ал жастардың көпшілігі кейбір мақал-мәтелді естіп жүргенмен, оның мағынасын жете түсінбейтін жағдайға жеткендігі өтірік емес. Осы орайда Ә.Қайдардың мақал-мәтелдер мағынасын мәдени-уәждемелік әдіспен түсіндірген әдістемесі бұл корпуста қамтылмаған мақалдарды түсіндіруде қолдануға болатын бірден-бір әдіс екендігін атап айтамыз.

Мақал-мәтелдер корпусында қамтылған мақалдар түсіндірмесінен үзінді берейік:

Кесте 18

Мақалдар түсіндірмесінен үзінді

Байлық байлық емес, бірлік – байлық	Байлық – өмірлік құндылық емес, өмірде байлықтан да жоғары тұратын құндылықтар (денсаулық, амандық, рухани байлық т.б.) бар. «Бірлік бар жерде тірлік бар» демекші, ынтымақты елде, ауызбіршілік бар жерде, жұмыла істелген істе нәтиже бар.
Ынтымақ жүрген жерде, Ырыс бірге жүреді.	Ынтымақтасып, бірлесіп атқарған іс берекелі болады. Қай істе болмасын ынтымақ, бірлік болса, ырыс-берекесінің болуы да заңды деген сөз.
Бір кісінікі – мақұл. Екі кісінікі – ақыл	Қандай да бір істі ақылдасып, кеңесіп шешкен абзал деген мағынада.

Программа іздеу кезінде табылған мақал-мәтелдер санын да көрсетеді.

Келешек зерттеулерімізде Мақал-мәтелдер корпусына салынған мақалдар мәтіндерінің жиілік сөздігі жасалатын болады. Бұл арқылы қазақ мақал-мәтелдерінде қай сөздің жиі қолданылатынын, қазақтардың мақал-мәтел шығаруда қай тақырыпты маңызды орынға қойғандығын т.б. ақпараттар алуға болады.

9.2 Белсенді қолданыстағы паремиологизмдер

Тілдің халық өмірінде атқаратын функциясы алуан түрлі. Солардың бірі – ұрпақ пен ұрпақ арасын жалғастыратын, өткен өмір мен бүгінгі болмысымызды, онымен болашақты жалғап жататын, ұлт тарихы мен менталитетіне, сана-сезіміне, өзіндік болмысына қажетті рухани құндылықтарды өз бойында сақтап, келесі ұрпаққа асыл мұра ретінде жеткізіп отыратын кумулятивтік қызметі. Ұлт даналығын араға талай ғасырлар салып, бүгінгі ұрпаққа нақыш-бояуын солғындатпай, мән-мағынасын бәсеңдетпей, көркемдігі мен бейнелілігін жоғалтпай, тұтас жеткізіп отырған тілдік бірліктердің бірі – мақал-мәтелдер. Қоғамдық өмірдегі түрлі оқиғаларды, бір халықтың бүкіл тұрмыс-тіршілігін, салт-санасын, тіпті тарихын, сан ғасырлық тәжірибесін көрсететін мақал-мәтелдер халықтың ауызша энциклопедиясы іспетті.

Дүниенің тілдік бейнесін жасайтын адам дейтін болсақ, адамның танымы, қарым-қатынасы арқылы тілді танимыз. Ендеше, тіл мен ойлаудың, ұлттық дүниетанымның тоғысуынан туындайтын күрделі құбылыстың бірі болатын мақал-мәтелдер қазақ халқының өмір тәжірибесінен, тіршілік танымынан түйіп түзілген ерекше лингвомәдени бірліктер. Этностының таным үрдісі мен рухани құндылықтарды қабылдау, зерделеу, пайымдау, тұжырымдау өрісі мақал-мәтелдерде кең көрініс тапқан.

Академик Ә.Қайдар «кез келген этностың (бүгінгі ұлыстың, халықтың не ұлттың) тілінде оның басынан өткен бүкіл өмірінің өрнегі жатыр. Халықтың, яғни этностың шын мәніндегі болмысы мен дүниетанымы оның тілінде ғана сақталады. Әрбір дәуірде өмірде қажет болған құрал-сайманның, қару-жарақтың, киер киім мен ішер тамақтың, тұрмыстық заттар мен салт-санаға, әдет-ғұрып, наным-сенімге, ойын-күлкі, той-томалаққа байланысты ұғымдардың аты-жөні, сыр-сипаты т.б. тек тіл фактілері ретінде ғана, яғни тілдегі сөздер мен сөз тіркестері, фразеологизмдер мен мақал-мәтелдер, аңыз-ертегілер арқылы ғана бізге жетуі мүмкін» деп пайымдаған [243, 35 б.]. Демек, мақал-мәтелдер – қазақ тілі сөз қорының ғасырлар бойы қалыптасқан тұрақты да құрамды бөлігі ғана емес, сонымен қатар ұлттық дүниетанымымыз бен менталитетіміздің тілдегі ең бір жарқын көрінісі.

Мақал-мәтелдердің тілімізде, өмір-тіршілігімізде, рухани-материалдық мәдениетімізде қолданылмайтын саласы жоқ. Сондықтан да оларды этностану, қоғамтану, табиғаттану, т.б. танымдық құбылыстардың тілдік негіздері деп қарауға болады. Оның өзіндік себептері де жоқ емес. Олардың ең бастысы тілдің танымдық қызметіне байланысты. Ал бұл тәсілдің түп-тамыры, сайып келгенде, тым әріден, көшпелі қауымның өмір-тіршілігіне сәйкес қалыптасқан тілдік дәстүрден бастау алғандығы сөзсіз.

Екі мыңнан астам қазақ мақал-мәтелдерінің табиғатын зерттеп және олардың түсіндірме сөздігін жасаған академик Ә.Қайдар: «Тілімізде қордаланған мыңдаған мақал-мәтелдердің мән-мағынасына зерделі көзбен қараған адам ең алдымен өз халқының басынан кешірген барша өмірін, салт-сана, қалыптасқан дәстүр, дүниетанымын елестете алады» дейді [272, 23 б.].

Мәселен, ғасырлар бойы мал шаруашылығымен айналысқан қазақтардың тілінде төрт түлік малға қатысты мақалдар көп: *Малды бақсаң, қойды бақ, май кетпейді шарадан; Мал өсірсең, қой өсір, өнімі оның көл-көсір; Қойың болмаса, байлықта ойың болмасын; Қой өрісі – қорадан; Қой шелді болмай, төлді болмайды; Күн батса, қойға соқпақ табылады; Қотыр ат соқыр атқа жақын жүреді; Атың барда желіп жүріп, жер таны, Асың барда беріп жүріп, ел таны.*

Мақал-мәтелдердің пайда болып, қалыптасуына ішкі/сыртқы факторлар әсер етеді. Солардың ең бастысы: адамның өмір сүріп жүрген табиғи ортаны, табиғат құбылыстарын танып-білуіне, еңбек процестерін, кәсіпшілікті меңгеруге, адамдар арасындағы қарым-қатынастардың қыр-сырын терең ұғынуға байланысты болды. Сондай-ақ табиғат құбылыстарының өзгеруі мен даму заңдылықтарын танып-білуден, тірлік-тіршілік көздерін меңгеруден, адамдар арасындағы сан алуан қарым-қатынастардың қыр-сырын терең ұғынудан

туындаған ой-тұжырымын, тәжірибе-тағылымын көп сөзбен ұзақ баяндаудан гөрі, тұжырымдап, қорытып, қысқа да нұсқа түрінде айтып жеткізу мақсатына байланысты.

Паремиологизмдер этностың этникалық сипаты, мәдениеті, мінез-құлқын, ішкі жан дүниесін, халықтың ғасырлар бойы көңілге түйгенін, сұрыптап жеткізеді. Қазақ тіл біліміндегі мақал-мәтелдер түрлі тақырыптарға топтастырылып, олардың қоғамдық өмірдегі түрлі оқиғалар мен құбылыстарды, адамның тұрмыс жайын, шаруашылық кәсібін, басқа да түрлі іс-әрекетін түгел қамтып, адамгершілік, тәлім-тәрбие, мінез-құлық жайлы даналық ой қорыту арқылы ұлттық дүниенің тілдік бейнесін жасаушы негізгі индикаторлардың бірі екені сөзсіз.

Соңғы кездері тіл білімінде адамның сана болмысын, оның айналасын, рухани қарым-қатынасын антропоэзектік және тілдегі адамдық фактор ретінде зерттеу белең алды. Халықтың элеуметтік-экономикалық өмірінің, оның рухани дүниесі мен материалдық мәдениетінің әдет-ғұрып, салт-дәстүр т.б. түсініктерінің қозғаушы күші, тірегі, діндегі адам болып табылады. Сондықтан да болар кез келген халықтың фразеологизмдері, оның ішінде мақал-мәтелдер адам ұғымымен байланысып жатады. Мұның бәрі ғаламның тілдік бейнесін жасаудағы ой иесінің үлес салмағын танытады.

Антропоэзектік ұстаным негізінде мақал-мәтелдер ауыспалы мағынада қалыптасып, тілдік қолданыста қарым-қатынас құралы ретінде жұмсалады. Мақал-мәтелдегі шындықтың бейнелі көрінісі өмірдің әртүрлі құбылыстарын эстетикалық бағалаумен байланысты. Эмоционалды заряд іс-әрекетке деген көзқарасты қалыптастырады, бұл мағынаның кеңеюіне әкеледі.

Қазақ мақал-мәтелдерінің барлығы дерлік ауыспалы мағынада қолданылады. Бұдан шығатын қорытынды мақал-мәтелдердің қалыптасуына негіз болған лексика-семантикалық шоғырдың алғашқы (номинативтік) мағынасының да болғандығын, яғни сол мағынаның абстракцияланып, ауыс мағынаға өтуіне қажетті ұлттық ұғым, ментальдық сана, логика-семантикалық мотив (уәж) т.б. заңдылығын көрсетеді [272, 36 б.]. Яғни паремиологизмдер уақыт өте келе әртүрлі лексика-семантикалық модификацияларға, мағыналық элементтерінің күңгірттенуіне ұшырайды.

Мақал-мәтелдер тұрақты тіркес санатына жатқанымен өзгермей, қасаң болып қатып қалған құбылыс емес. Олар да жанды тілдің, ой толғамның өзгермелі ерекшеліктері мен заңдылықтарына сәйкес тұлғасы мен мағынасы құбылып, әр алуан модификациялық күйге ұшырап отырады. Сондықтан тілімізде жаңа мақал-мәтелдер пайда болады, ескілері ұмытылады.

Біз зерттеу жұмысымызда белсенді қолданыстағы мақал-мәтелдерді анықтап, олардың қазіргі семантикасын анықтауды, мәдени-семантикалық мазмұнын ашуды мақсат еттік. Алдымен, белсенді қолданыстағы паремиологизмдерді анықтау үшін 18-35 жас аралығындағы, қазақ тілді аудиторияға «Қандай мақал-мәтелді жиі қолданасыз», «жиі еститін мақал-мәтеліңіз қандай, ол мақалдың мәнін түсіндіріп беріңіз» деген сұраулар қойылды. 46 респондент қатысып, мына мақал-мәтелдерді күнделікті өмірде жиі қолданамыз деп көрсетті: *Еңбек етсең, емерсің; Білгенім бір тоғыз, білмегенім тоқсан тоғыз; Балық*

басынан шіриді; Ит жоқта шошқа үреді; Сенген қойым сен болсаң...; Отан отбасынан басталады; Шешесіне қарап қызын ал; Тәрбие тал бесіктен; Тегі жарымағанның тілеуі жаман; Еңбек етсең ерінбей, Тояды қарның тіленбей; Еңбек бәрін жеңбек; Күріштің арқасында күрмек су ішер; Әзілің жарасса, атаңмен ойна; Білім – инемен құдық қазғандай; Басы істемегеннің екі аяғына зор келеді; Балалы үйдің ұрлығы жатпас; Балалы үй – базар, Баласыз үй – құ мазар; Қызым саған айтам, келінім сен тыңда; Қатынды жастан, баланы бастан; Той десе құ бас домалайды; Ет етке, сорпа бетке; Тойға барсаң, тойып бар; Көппен көрген ұлы той; Тойдың болғанынан, боладысы қызық; Жаманнан жарты аяқ ас қалады; Шақырылмаған қонақ сыпырылмаған жерге отырады; Бала белде, қатын жолда; Тең теңімен, тезек қабымен; Көп түкірсе – көл; Алыстағы ағайыннан, жақындағы көршің артық; Көп жасаған білмейді, көпті көрген біледі; Молданың істегенін істе ме, айтқанын істе; Аңқау елге арамза молда; Алдыңғы арба қайда жүрсе, соңғысы сонда барады; Ерте тұрған еркектің ырысы артық, ерте тұрған әйелдің бір ісі артық; Көп тыңда, аз сөйле; Сыйласаң сыйлы боласың; Аузы қисық болса да байдың баласы сөйлесін; Қыстың қамын жаз ойла; Адамның күні адаммен; Болар елдің баласы бірін бірі батыр дер; Бір көрген – біліс, Екі көрген – таныс; Есектің артын жусаң да мал тап; Жақсылық қылсаң, бүтін қыл; Еріншектің «ертеңі» таусылмайды; Дос жылатып айтады, Дұшпан күлдіріп айтады; Көзден кетсе, көңілден кетеді; Сабыр түбі – сары алтын; Сақтанғанды Құдай сақтар; Аузы күйген үрлеп ішер; Қасқырдың аузы жесе де қан, жемесе де қан; Барымен базар; Ағын суда арамдық жоқ; Жылы-жылы сөйлесең, жылан інінен шығар; Шығын шықпай, кіріс кірмейді; Қайтып келер есіңді қатты жаппа; Сыйға сый, сыраға бал; Атасы құрдастың баласы құрдас; Басқа бәле тілден; Жалғыздық Құдайға гана жарасқан; Алыстан арбалағаниша, жақыннан дорбала.

Ал өзіңіз жиі қолданатын мақал-мәтелдердің мазмұнын ашып көрсетіңіз деген сауалға төмендегідей жауап берді:

Адам(нын) күні адаммен, (Дос болма топас наданмен). «Адам адаммен жасасады» демекші, күллі адамдармен жақсы аралассаң да, ішінен дос деген атқа лайық адамды тандай білу. Сол себепті, көп достың ішінен адал жанашыр досты дұрыс тандай білу керек.

Дос жылатып айтады, Дұшпан күлдіріп айтады. Дос – адамның жалпы өмірінде елеулі орын алатын, өзіндей көретін аяр адам. Сондықтан мақалда адал дос саған шындықты ащы болса да батырып айтады, ал сенің дұшпаның шындық ащы болса да оны тәтті қылып, үстіңнен күліп айтатындығын айтып тұр.

Сабыр түбі сары алтын, (Сарғайған жетер мұратқа). Сабырлылық – адам бойындағы ең ізгі қасиет. Сабырлы адам дйттеген мақсатына бірдеммен жетеді. Сондықтан адам баласы неғұрлым сабырлы болған сайын соғұрлым ол адам тілегенін еселеп алады.

Барымен базар, (Жоқты қайдан табар). Қолдағы бар дүниемен қанағат ете білу керек. Мақал қолдағы барға қанағат, жоқ болса алдағы уақытта бола жатар дегенді меңзеп тұр.

Ағын суда арамдық жоқ. Ағын су – жер бетіндегі жоғарыдан төмен қарай ағып жататын су. Мақал ағын су үнемі тоқтаусыз бір арнамен ағып жа-

нарып отыратындықтан, оның бойында арамдық тұрмайды. Сол секілді кейбір адамдар ағын су сияқты бүкіл жан-дүниесін ақтарып, ағынан жарылады. Бұндай адамдардың бойынан арамдық байқалмайды.

Әлін білмеген әлек. Әрбір адам өзінің күшіне қарап жұмыс істеуі керек. Мақалда өзінің шамасын білмеген адам өзінің жағдайын өзі қиындататындығы айтылып тұр.

Шығын шықпай, кіріс кірмейді. «Шығыс бар жерде кіріс бар» демекші, бір істі бастамас бұрын алдымен қолындағы барын берсе, кейіннен сол бергені еселеніп қайтады. Мақал басында қиналып шығын шыққан кезде артынан кірістің болатынын меңзеп тұр.

Есектің артын жу да мал тап. Мақал еш еңбектенбей тек жатып ішер болып, жалқауланып кеткен адамға ұнамсыз жұмыс болса да жасау керектігін айтып тұр.

Қайтып келер есіңді қатты жаппа. Әрбір адам түрлі эмоцияға беріліп, кейде шектен шығып жатады. Мақал қандай жағдай болса да қайта келетін немесе бет көрісетін адамдармен сыйластықты сақтап, қайта кіретін есіңді қатты жаппау керектігін айтып тұр.

Байқағандарыңыздай, мақалдардың барлығының дерлік бастапқы мағынасы, мәні сақталған, алайда айта кететін жайт, мақалдардың көпшілігі білім, әлеуметтік жағдай, тәрбие, достық, той, қонақ, т.б. туралы болды. «Қазақ халық мақал-мәтелдерінің басым көпшілігі қазақ шаруасының негізгі күнелткіші болған малшылық пен егіншілікті, көмекші кәсіптері – аңшылық пен балықшылықты бейнелейді. Онда шаруаның және диханның мінез-құлқы, тұрмыс-салты, ұғым-сенімі тапқыр ой, көркем тілмен жан-жақты сипатталған; шаруашылық жүргізудегі, қарым-қатынас жасаудағы тәжірибелері мен көрген-білгендері қорытылғандығын» [273] айтады. Бұл – мақал-мәтелдер өмір тіршілігінің айнасы екендігінің көрсеткіші. Мақал-мәтелдердің пайда болуына қандай да бір құбылыстың, күнделікті көзбен көріп, құлақпен естіп, жүрекпен сезініп жүрген оқиғалар әсер ететінін ескерсек, қазіргі қазақ қоғамында белсенді қолданысқа ие паремнологизмдердің арасында көшпелі қазақтың тұрмысында ерекше орын алған төрт түлік мал жөніндегі, көші-қон жөніндегі, қару-жараққа қатысты т.б. осы күнгі тіршілігімізде жоқ, мақал-мәтелдердің пассив қорға қарай өтіп бара жатқандығы байқалды.

Сондай-ақ бұл мақал-мәтелдер бастапқы құрылымы сан жағынан да, сапа жағынан да тарылып, ықшамдалып қолдануы кездесті (*Шешесін көріп қызын ал, аяғын көріп асын іш; Сабыр түбі – сары алтын, Сарғайған жетер мұратқа, Сабырсыз қалар ұятқа; Сенген қойым сен болса, күйсегеніңе болайын; Адамның күні адаммен, Дос болма топас наданмен; Барымен базар, Жоқты қайдан табар т.б.*).

Сонымен, қорыта келе, «мәдениетіміз, ой-санамыз қандай болған, мынау ақиқат дүниені біз қалай таныдық, қалай жүйеледік, оны қалай көрсете алдық?» деген сұрақтарға толық жауап бере алатын, тоқсан ауыз сөздің тобықтай түйінін беретін паремнологизмдер үлкендердің жас ұрпаққа айтатын өнеге-өсиеті, ғибрат сөздері, ақыл-кеңесі ғана емес, сол ұлттың өзі туралы кең ақпарат беретін дереккөз болып табылады.

10 ЖАРНАМА ІШКОРПУСЫ

10.1 «Қазақ» газетіндегі жарнама мәтіндерінің құрылымы

Жарнама сауданың дамуына орай тауар, ұсынылатын қызмет түрлерінің ерекшеліктері туралы ақпарат беретін, артықшылықтарын танытатын, осыған орай тауарға, қызметке деген сұранысты көбейтетін хабарлама беру қызметінде пайда болды. Өркениеттік дамудың бір артықшылығы мәдени қауымдастықтың, аралас-құраластық пен сан қырлы байланыстың, коммуникацияның жоғары даму сатысына ие болғандық екені белгілі. Сол тұрғыда жарнама да өз қызметін артығымен атқаратын қауымдасудың бір тетігі болып қалыптасты.

Жарнама алғашында әлеуметтік қатынастарды реттеуші мақсатында пайда болды. Әлеуметтік жіктер мен топтарды, өзім мен өзгені айыру үшін жарнаманың алғашқы нышандары қызмет етті.

Қазіргі таңда жарнама графикалық кеңістіктегі коммуникацияның қоғам назарын аударатын басты құралына айналды. Заманның техникалық жетістігімен байланысты жарнама мәтіндері де түрлі форматқа ауыстырылды. Көшедегі (сыртқы) жарнама, интернеттегі жарнама, БАҚ-тағы (газет, журнал, теле-радио) жарнама, телефондағы жарнама (YouTube, Facebook, Instagram, Tiktok) түрлері көп.

Жарнама әлеумет тарапынан жағымсыз қабылданса да, қоғамдық-әлеуметтік қатынасты анықтаушы, дамытушы күшке айналды [274].

Қазіргі қазақ қоғамы жарнаманы негатив түрде қабылдаудан бейтарап түрде қабылдайтын күйге жетті. Сондықтан жарнама түрлері алуан түрлі болып, кеңінен қызмет етіп келеді.

Қазақ даласында жарнама жасау тарихы 1913-1918 жж. Алаш зиялылары ұйымдастырып, шығарған алғашқы ұлттық «Қазақ» газетінен бастау алады. Жарнама нышандары «Қазақ» басылымына дейінгі мерзімді баспасөздерде де болды, бірақ жарнама беру арқылы басылымға демеуші болу тәжірибесі, жарнама жариялау арқылы газеттің тиражын көбейту әдісі «Қазақ» газетінен бастау алған сыңайлы. Сондықтан қазақ тіліндегі жарнама тарихын зерттеуді ұлттық тіліміздің негізін салған алғашқы ақпарат құралы – «Қазақ» газетіндегі жарнама мәтінінің құрылымын, прагматикасын, тілі мен стилін зерттеуден бастаудың өзектілігі бар.

Зерттеу барысында сипаттау, салыстырмалы талдау, жүйелеу және тұжырымдау әдістері қолданылды.

Зерттеу жүргізу үшін алдымен 1913 жылғы «Қазақ» газетіндегі жарнама мәтіндері жарияланған нөмірлер жиналып, сканері жасалды. Төте жазумен жазылған мәтіндердің кирил жазуына графикалық проекциясы жасалды. Графикалық проекция деген ұғым алғаш рет А.Байтұрсынұлы атындағы Тіл білімі институтының директоры, ф.ғ.к. Анар Фазылжанның негіздеуі бойынша «Тіл – құралдың» 1-кітабының 1914, 1924, 1927 жылғы басылымдарын кирил графикасына проекциялау кезінде қолданылды. Үш басылым арасындағы жазу айырмашылықтарын, графиканың таңбалану ерекшеліктерін көру үшін үш басылымның транслитерациясы немесе транскрипциясы емес, графикалық

проекциясы (графикалық бейнесі) кирил әрпінде бейнеленген алғашқы ғылыми тәжірибе болатын [275]. Осы бағытта «Қазақ» газетіндегі алғашқы жарнама мәтіндерінің де төте жазудағы берілу форматын түпнұсқа қалпында көру үшін графикалық проекциясы жасалды. Жарнамалардың графикалық проекциясын жасаған осы мақала авторларының бірі – филология ғылымдарының кандидаты, «Ахметтану» бөлімінің жетекші ғылыми қызметкері Гүлфар Мамырбек.

Зерттеу нысанына 1913 жылғы «Қазақ» газетінің №7-40 аралығындағы жалпы саны 40-тан аса жарнама мәтіні алынды. Жарнама мәтіндері мазмұнына қарай тақырыптық топтарға бөлінді.

Жалпы қазақ тіліндегі жарнама тарихына, тіліне қатысты жарық көрген ғылыми зерттеулер қатарына С.Асанбаеваның «Жарнама саласында мемлекеттік тілді қолданудың социоллингвистикалық өзекті мәселелері» (1999), Ғ.Солтанбекованың «Қазақ жарнамалары: тілдік сипаты және олардың ықпал ету қызметі» (2001), Л.Бугенованың «Рекламные тексты на казахском и русском языках в межкультурной парадигме» (2007), М.Увайсованың «Жарнама тілі: лингвопрагматикалық аспект» (2010), М.Шындалиеваның «БАҚ-тағы жарнама технологиясы» (2012), А.Байдуллаеваның «Құрылымы әртүрлі тілдердегі жарнама слоганы» (2016) атты еңбектерін жатқызуға болады. Мақала жазу барысында жарнамаға қатысты ғылыми-теориялық еңбектердегі тұжырымдар назарға алынды.

Жарнама атауы парсы тіліндегі *жар* және *наме* деген екі сөздің бірігуінен жасалған. *Жар* сөзі «елге хабарлау, білдіру, жариялау» деген мағынаны білдіреді. Қазіргі қазақ тілінде бұл сөз *жар салу*, *жар қылу* тіркестерінде сақталған, ал ескі қазақ жазба тілінде *жар шақыру*, *жарламау* (жарияламау, білдірмеу), *жаршы* (хабарлаушы) сөздерінің құрамында актив қолданыста болды. *Наме* сөзі «жазба, жазылған нәрсе, жазылған хат, шығарма» деген ұғымды білдіреді. Бұл екі сөз қазақ тілінің дыбыстық заңдылығына бейімделіп, *жарнама* түрінде орнықты және тек жазылған хабарламаға ғана қатысты айтылмай, «кұлақтандыру» мәнінде қолданылды.

«Қазақ» газетінде *жарнама* сөзі дефис арқылы берілген: *жар-нама*. Және ондағы «жар қылып білдіреміз» деген сөйлемнен *жарнаманың* елге жар салып хабарлау ұғымында қолданылып тұрғаны анық аңғарылады:

Жар-нама

Ардақты оқушыларымыз!

Йұлдің басында 5 айға жазылған оқушыларымыздың газета алатын мөлшерлі уақыты толды. Сонда да оларға йул ақырынаша газета жібере тұрамыз. Сол секілді 6 айға жазылған оқушыларымызда да газета алатын мөлшерлі уақыты йул ақырында толады. Мөлшерлі уақыт толған соң газета жіберілмейді. Сондықтан жар қылып білдіреміз: газетамыз тоқталмай барып тұрсын десеңіз, күні бұрын кесімді жазылу ақшасын жібересіз.

«Қазақ» басқармасы» («Қазақ» газеті, №21, 1913)

Жар-нама

Феврал басынан газетамызга жазылған алты айлық надписиштеріміздің мөлшерлі уақыты осы 24 нші номырмен толды. Енді оларға газета жіберілмейді.

(«Қазақ» газеті, №24, 1913)

«Қазақ» газетіндегі жарнамалар, мәтінде айтылғандай, жар салу, айту, хабарлау, құлақтандыру мақсатында жарияланады. Бұл алғашқы жарнамалар «Қазақ» газетінің 1913 жылғы 21-ші, 24-ші нөмірлеріне «Жар-нама» деген атаумен шығады. Және бірден қазақ тілінің үндесім заңына сәйкес бірыңғай жуан үндесіммен жар-нама деген толыққанды атау түрінде пайда болады. Шындығында, бұл сөз жар-наме деп түпнұсқа тілге жақын жазылуы да мүмкін еді.

Газет шығарушылардың өз сөзімен айтқанда негізі мақсат: «Жұрттың пайдасына көз болу, қазақ арасына ғылым, өнер жайылуына көз болу, қазақ арасына ғылым, өнер жайылуына басшылық ету, басқа жұрттардың халінен хабар беріп, таныстыру. Сол мақсаттарды орнына келтіру үшін зақұндерді, хакімдердің бұйрық-жарлықтарын білдіріп тұру. Мемлекеттік Дума һәм Мемлекеттік кеңес жұмыстары турасында жете хабар беру, ішкі һәм сыртқы хабарларды жазып тұру, қазақтың бұрынғы және бүгінгі жайын жазу, күнелту, кәсіп, жер-су, егін-таран, мал шаруашылығы, оқу оқыту, мектеб, медресе, школа, ғылым, өнер, әдебиет турасында жөн көрсету, адам һәм мал дәрігерлігі жайынан жазу» болуының өзі арлы-берлі хабарларды бірінен-екіншісіне жеткізу болған деп келтіреді, «Қазақ» газетінің тілі мен стилін зерттеген профессор [276, 51 б.].

«Қазақ» шығарушылары баспахана сатып алуды о бастан көздегенін, оқырмандардың, қалталы адамдардың, Қытайдағы қазақтардың көмегімен, құрған серіктестігінің арқасында 40 мың сомға қазақ даласындағы алғашқы баспахана осылай ашылғанын айтады [277, 128 б.].

Осыған орай алғашқы қазақ ұлттық баспасөзінің пайда болуына, үздіксіз жариялануына жарнамалардың жариялануы айтарлықтай жәрдем еткен деп ойлаймыз. Жарнама беруші өзі туралы ақпаратпен жұртшылықты құлақтандыру үшін «Қазақ» газетінің оқылымыдылығын пайдаланғаны үшін газетке жарна құйып тұруы әбден кәдік.

Профессор Б.Момынова «сол кездегі мерзімді басылымдардың бәрінің қазақ халқы үшін қызметі әмбебап, энциклопедиялық сипатта болды. Бірақ бәріне ортақ бір қызмет түрі болды, ол – ағартушылық қызмет еді. «Қазақ» газеті осы қызметінде өзгелерден оқ бойы озық басылым еді. Өйткені газет А.Байтұрсынұлы түрлеген төте жазуды іс жүзінде алғаш пайдаланып, тәжірибеден өткізген, жазудың жүйеге түсіп, реттелуіне зор үлес қосқан, усул-жәдид қозғалысын қазақ даласына кең жайған газет болды» деп айрықша қызметін даралап көрсетеді [276, 51 б.].

Аталған газет жарнама қызметін ағартушылық мақсатта кеңінен қолданған газет болды. А.Байтұрсынұлы төте жазудың, емле түзудің реттелуіне тірек

болған еңбектерін өзі бас болып шығарған осы «Қазақ» газетінде жарнамалап, хабардар етіп отырғанын айтады ғалым: «1913 жылғы 9-нөмірінде (6 апрель) «Жақын арада басылып шыққан һәм шығатын кітаптар» тақырыпты шағын хабар бар (осындай хабарлар әр жылдары уақ-уақ жарияланып отырған). Онда: «Оқу құралы» есімінде үш кітапша шығады. Біреуі қазақ тілінде қандай дыбыстар бар, ол дыбыстар қандай хәріппен, қалай жазылмақшы һәм усул саутиә (сотиве) жолымен оқытуды қалай бастау афдал болатын жайын сөйлейтін «Баяншы» кітап. Бұл басылып шықты. Бағасы пошта ақысы 5 т., пошта ақысымен 7 т. Онан басқа екі кітапшаның біреуі – қазақ тіліне жазылатын хәріптерді үйрететін әліпби. Екіншісі – қазақтан басқа тілдерге, мысалы: басқа түрки, араби, фарси сөздеріне жазылатын хәріптерді үйретуге шығарған екінші кітап. Бұл соңғы кітапта әріптердің бәрін білгеннен кейін балалар оқуға төселуіне ыңғайлап жазған қазақша мақалдар, жұмбақтар, баталар, өлеңдер, ұсақ-ұсақ қарасөздер бар. Бұлар тез арада басылып шығады. Адрес мынау: Оренбург, Редакция газеті «Казак». А. Байтурсунову» [276, 52 б.].

Төмендегі сурет (25-сурет) «Қазақ» газетінің соңғы бетінде берілген жарнама форматын көрсетеді. Бұнда бірыңғай мәтін түрінде емес, оқырманның назарын аударатын майлы бояумен қоршауға алынып, тақырып та майлы бояумен, ірі қаріппен беріліп, сурет, сызбалармен безендіріліп отырғанын көре аламыз.



Сурет 25. «Қазақ» газетінің соңғы бетінде берілген жарнама форматы

«Қазақ» газетінде әртүрлі бағыттағы жарнамалар жарияланып отырған. Оларды бірнеше топқа бөлуге болады.

- 1) Әлеуметке пайдалы ақпарат тарату үшін берілген хабарландыру сипатындағы жарнамалар;
- 2) Тауардың сұранысын арттыру мақсатында берілген жарнамалар;

3) Көрсетілетін қызметке қызығушылық тудыруға бағытталған жарнамалар.

Мәселен бірінші топқа жататын жарнамаларға «Қазақ» газетіне, басқа да журналдарға жазылу ақысы мен мақала жариялау ақысын көрсеткен мәтінді, газетке қандай материалдар қабылданатыны туралы хабарды, жаңадан шыққан кітаптар туралы ақпаратты жатқызуға болады. Яғни бұлар жарнаманың негізгі, бастапқы қызметіне сәйкес, елге жар салып, құлақтандыру мақсатында жарияланған. Әрі бұл ақпараттар бір рет қана жариялана салмаған. Мысалы, «Қазақ» газеті туралы мына хабарлама әр нөмір сайын шығып отырған:

«Жазылу бағасы:

Жылға 3 сом, жарты жылға 1 сом 75 тиын, үш айға 1 сом. Номыр бағасы Орынбурға 5 тиын, басқа жерлерде 7 тиын.

Иглан бағасы:

Бір рет басуға 1 ниі бетте петит жолы 20 тиын, 4 ниі бетте 15 тиын. Бір реттен артық басу бағасы келісуінше.

Алдырушыларға адресі мынау: Оренбургь, Гришковская ул. Д.Димитриевой.

Редакция газеты «КАЗАК»

**Газетамызға қай тіл мен болса да жазылған сөздер,
хабарлар қабул алынады**

Газетамыздың көп мақсұдының бірі: басқадан да алып, өзімізде барын да жиып қазақ тілінің бетін жөндеп жолын ашап, сондықтан қазақтың мақалдары, тақпақ сөздері, жұмбақтары, билердің, шешендердің айтқан сөздері, басылмаған өлеңдер, қиссалар – бәрі де қабул алынады.

Басылмай қалған сөздер қайтарылмайды.

*Адрес: Орынборг, Гришкоуски орамда, Дмитриива үйінде,
«Қазақ» басқармасы» («Қазақ» газеті, №7, 1913)*

Жоғарыда берілген жарнамалар тілінде *иглан* (жариялау), *петит* (типografíaлық шрифт), *орам* (көше) тәрізді сол кезеңнің тілдік қолданысына тән кірме сөздер кездеседі. 1913 жылы жарияланған жарнамаларда қазіргі тілде қолданылмайтын араб, парсы сөздері біршама, әрі көптеген терминдер араб, парсы сөздері арқылы жасалған:

Мухаррар – редактор

Иқтисади – экономикалық

Милли – ұлттық

Матбуғат – баспасөз

Ибтидаи – бастауыш

Рашиди – ересек

Шарика – серіктестік

Шаһадат нама – куәлік қағаз

Мысқал – грамм

Қадар – кішкентай, аз

Исмлі – атты, аталатын

Усул – тәсіл, жол

Қирағат – мәнерлеп, мақаммен оқу

Асла – асылы, т.б.

«Қазақ» газеті Ахмет Байтұрсынұлы түзген төте жазудың тәжірибе алаңы болып, ұлттық емле, ұлттық әдеби тілдің нормаларын қалыптастырушы қызмет атқарғанмен төте жазу бірден жәдид жазуынан қол үзіп кете алмады. Терминдер араб тіліндегі нұсқасымен қолданылуы және жазылуы жазу тәжірибесіндегі дәстүрді бұзбау принципінен шыққаны сөзсіз. Және зерттеуге алынып отырған газеттің алғашқы нөмірлері болып табылатынын ескерсек, жоғарыдағы мысалдардың қолданылуы заңды. Сонымен бірге жарнамаларда бірқатар терминдер орыс тіліндегі нұсқасымен алынған:

Акружни (окружной) – аудандық

Присажни (присяжный) – алқаби

Углауни (уголовный) – қылмыстық

Граждански (гражданский) – азаматтық

Администратиуни (административный) – әкімшілік

Тұуаристуа (товарищество) – серіктестік

Плошиат (площадь) – алаң

Басна (басня) – мысал, тәмсіл

Торгоуи (торговый) – сауда

Наложни платеж (наложенный платеж) – үстеме төлем

Қобанқа (кубанка) – бидайдың сорты, қатты бидай

Земстуа (земство) – өзін-өзі басқару органы

Затадке (задаток) – кепіл

Прауа (право) – құқық

Судебни палата (судебная палата) – сот палатасы, т.б.

Өйткені бұл қатар қазақ даласына енді еніп келе жатқан қоғамдық-саяси, іскери, экономикалық қатынастарда жиі қолданылып келе жатқан сөздер еді.

Жоғарыда берілген жарнама мәтіндері практикалық транскрипция негізінде қазіргі емлеге сәйкес ұсынылды. Алайда ХХ ғасырдың басындағы жазу нормасы басқаша болғаны мәлім. Мәселен, араб, парсы сөздері сол кезеңдегі жазу емлесіне сәйкес цитаталық принциппен, түпнұсқа тілдегі жазылымы бойынша берілген: *бага* сөзі *баһа*, *емле* сөзі *имла*, *қызмет* сөзі *хизмат*, *зейін* сөзі *зеһн*, сондай-ақ *китаб*, *дафтар*, *сыныф*, *мактаб*, *асбаб*, *китабхана*, *харф*, *фарси*, *файда* т.б. Тек кірме сөздер ғана емес, төл сөздердің жазылымы да сол кезеңнің емлесіне сәйкес таңбаланған. Оны жарнаманың практикалық транскрипциясы мен графикалық проекциясын салыстыру арқылы айқын аңғаруға болады (19-кесте):

Жарнама мәтіндерінің проекциясынан үзінді

Графикалық проекциясы	Практикалық транскрипциясы
«Қазақ» газетінің 1913 жылғы 15-ші нөміріне шыққан жарнама мәтіні	
Баслб шқды «АЗАМАТ» исмли қазақша һәр түрлі ғибратлы өлеңдер. Шғарушы Мир Иақуб Дулатф. Кітап баһасы 20 тин, көбдеб алдрушыларға 30% кемтледі.	Басылып шықты «АЗАМАТ» есімді қазақша әр түрлі ғибратты өлеңдер. Шығарушы Міржақып Дулатов. Кітап бағасы 20 тиын, көптеп алдырушыларға 30% кемітіледі.
«Қазақ» газетінің 1913 жылғы 18-ші нөміріне шыққан жарнама мәтіні	
Екнші рет баслб шқды ҚРҚ МСАЛ Ахмед Байтұрсынфдң қазақша өлең мен мәшһүр Крылоф басналарынан тәржме айтб шғарған кітабы. Бұл кітабдң тлі жеңл, баслууы әдеме, имласы «қазақ» имласы. Баһасы 25, пошта арқлы 30 тин.	Екінші рет басылып шықты ҚЫРЫҚ МЫСАЛ Ахмет Байтұрсынфтың қазақша өлең мен мәшһүр Крылов басняларынан тәржіме айттып шығарған кітабы. Бұл кітаптың тілі жеңіл, басылуы әдемі, емлесі «қазақ» емлесі. Бағасы 25, пошта арқылы 30 тиын.

Жарнамалардағы келесі бір ерекшелік, негізгі мәтін араб жазуымен, ал мекен-жай кирил жазуымен де қоса көрсетіліп отырған. Мекен-жайдың үнемі кирилмен де беріліп отыруы сол кезеңдегі пошта қызметінің талабымен байланысты болған тәрізді.

«Қазақ» газетінің алғашқы нөмірлерінде тауардың сұранысын арттыру мақсатында берілген жарнамалар қатары біршама. Оларға кітап дүкені, дәрі, тырма, шөп жинайтын машина, сүт тартатын машина, іс тігетін машинаны насихаттаған, соларды сатып алуға үгіттеген, яғни коммерциялық бағыттағы жарнамалар жатады. Бұл жарнамалардың құрылымындағы ең басты ерекшелік – тұтынушы назарын аударатындай слоган компоненттерінің болуы. Зерттеушілер слоганды сәтті қолданған жағдайда «Ол жарқын жарнама бейнесін жасауға себепші болып, жарнаманың ықтимал сатып алушыға жасайтын ықпалын арттырады» – деп тұжырымдайды. «Слоган» ұғымы галл тілінен аударғанда «айқай, ұран» дегенді білдіреді» [278, 40 б.]. Яғни слоган мақсаты – тұтынушы назарын аудару, сатып алуға ынталандыру, тұтынушы санасына әсер ету. «Қазақ» газетіндегі коммерциялық бағыттағы жарнамаларда тауардың сапасына көңіл аудартып қана қоймай, тұтынушының сезіміне де әсер ететін фразалар қолданылғанын аңғаруға болады. Мұны төмендегі шөп жинайтын машинаға қатысты жарияланған жарнамалардан анық көре аламыз:

«Шөп машинасы

Жер шаруасына керек машина сатып аларда ең әуелі беріктігін көздеу керек. Жұмыс қылғанда машина тоқталмайтын боларға керек һәм қосымша бөлек-салақ нәрселері қымбат болмасқа керек. Осы айтылған шартқа тура келетін «Геркулес» деген машина.

(«Қазақ» газеті, №15, 1913)

Шөп жиятын машина

Машинаның заты, құрылған нәрселері берік, асыл болса көп қызмет етеді. Мінеки, біз «Геркулес» дейтін шөп жиятын машинамызды істеткенде айтылмыс шарттарды орындауға аса зейін салдық. Сондықтан бұл машинамыздан машина аспайды.

(«Қазақ» газеті, №18, 1913).

Шөб машинасы

Бұл күнде жұмысшы қол қымбат. Шаруаға жұмысшы қолдың көбіне сүйенбейтін машиналар пайдалы. Ондай болатын жеңіл һәм берік, бөлек салағы қымбат түспейтін машиналар. Бұл шартқа тура келетін «Геркулес» ден артық машина жоқ.

(«Қазақ» газеті, №19, 1913).

Машина тырма

Шөп бітік болғанмен жұмысшы қол қымбат болса, пайда бере алмайды. Сондықтан жиюшы машинаның жақсысын алу керек. Ол «Геркулес» тырмасы.

(«Қазақ» газеті, №20, 1913).

Бір ғана «Геркулес» машинасына қатысты жоғарыдағы жарнамаларда слоган компоненттері адам эмоциясына әсер ете алатындай құбылып отырған. Мәселен, «беріктігін көздеу керек», «машина тоқталмайтын боларға керек», «нәрселері қымбат болмасқа керек», «айтылған шартқа тура келетін «Геркулес», «Геркулес» ден артық машина жоқ», «бұл машинамыздан машина аспайды», «Жиюшы машинаның жақсысын алу керек. Ол «Геркулес» деген фразалар тауар сапасының өте жоғары екеніне сендіре алады. Сондай-ақ төмендегі жарнамада да «Сібірдегі алтын кәсібі май кәсібінен кейін қалды», «май кәсібінің озуына себепкер болған «Алфа лавал», «айдай әлемге паш болған «Алфа лавал» тәрізді слогандар қолданылған:

«Алфа-нобел»

Тұуаристуасы (серіктесі)

Жылына отыз миллион сомдық мың жеті жүз алтын шығаратын Сібірдегі алтын кәсібі жылына төрт миллион пұт сүттен май істеп елу миллион сом шығаратын май кәсібінен кейін қалды. Алтын кәсібінен май кәсібінің озуына себепкер болған, айдай әлемге паш болған «Алфа лавал» деген сүттен май айыратын машина. Өзгеден мұның артықтығы сүтке май жібермейді. Өзі берік, тұтынуға оңай. Сібір жерінде май айыратын машинаның жүзден сексені «Алфа-лавал» машинасы. Сүттен май айыратын заводтарға керек құралдың бәрі «Алфа-нобел» де табылады.

(«Қазақ» газеті, №17, 1913)

«Қазақ» газетіндегі жарнамаларда слогандар өнім туралы жағымды көзқарас қалыптасыру және оған қызығушылық тудыру мақсатында қолданылған. «Слоган – жарнама мәтініндегі прагматикалық қызметі күшті тілдік құ-

рылымның бірі» [279, 23 б.]. Демек, слоганның бұл қызметі «Қазақ» газетіндегі жарнамаларда ерекше назарға алынған дей аламыз.

Жарнаманың негізгі ақпараттық бөлігі 3, асса 4 сөйлемнен құралады. Алғашқы сөйлемдер зат туралы білімге, тәжірибеге сүйенген негіздеме себептен тұрады. Соңғы бір ғана сөйлемде тауардың аты аталады, мысалы, «осы айтылған шартқа тура келетін «Геркулес» деген машина». «Бұл шартқа тура келетін «Геркулес» ден артық машина жоқ». «Сондықтан бұл машинамыздан машина аспайды». «Ол «Геркулес» тырмасы». «Сүттен май айыратын заводтарға керек құралдың бәрі «Алфа-нобел» де табылады» деген сөйлемдер жарнама ақпаратын аяқтайды. Ал алдыңғы сөйлемдер қажет затты таңдаудың оңды жолдарын түсіндіреді, себепін айтады; былайша айтқанда тыңдаушыға не оқырманға жұмбақ жасырады. Тыңдаушы негіздемені оқып, жөн сөз екен деген ойға тірелгенде, жұмбақтың сырын ашып, жарнаманың соңына жетуге асығады. Соңы бір ғана сөйлеммен, «тақ» етіп, жауап түрінде айтылады. Ары қарай сөйлем қосып, ұзартылмайды. Жұмбақтың жауабы нақты. Бұл жарнамалардан кейін затқа қызықпау, қайда сатылатынын іздемеу, бағасын білу үшін хабарласпау мүмкін емес. Жарнама прагматикасы осылай жасалған. Яғни бүгінгі жарнама мәтіндеріндегідей нысанның атауы бірнеше қайталанып, бұл затқа не мақсатпен назар аудару керектігі айтылмай, нысанның өзгелерлен артықшылығы көрсетілмей мәнсіз қайталана бермейді. Әр сөйлем мен сөзде мән, мағына бар.

«Қазақ» газетіндегі жарнамалардың келесі бір тобы – ақылы қызмет түрлерін ұсыну, яғни көрсетілетін қызметке қызығушылық тудыруға бағытталған жарнамалар. Оған адвокаттық, орысша үйрету, машинкаға теру қызметтері ұсынылған жарнамаларды жатқызуға болады. Мәселен:

«Орынбор акружни (окружной) соттың қарамағындағы присажни (присяжный) паернни (поверенный) (адуакат) Илия Михайлович Алексеев ізденуге алады:

Углауни (уголовный) (бас кетерлік), граждански (мал өндрерлк), админстратиуни (админстративный) (өкмет) істерді.

Жұмысы бар адамдармен сөйлесетін сағаттары:

9 дан 11 ге шейін (ертемен)

6 дан 8 ге шейін (кешпен)

Байрам күндерде 12 ден 2 ге шейін («Қазақ» газеті, №7, 1913).

Г.Загуменова Орысша жазуларды машинамен көшіріп басып беруді қабыл етеді. Қымбат алмайды» («Қазақ» газеті, №16, 1913).

Орысша оқытамын

Жаз айларында орысша оқушыларды қабыл етемін. Ауыл орыны бек көңілді: орман өзен жанында. Поштасы да жақын («Қазақ» газеті, №15, 1913) .

Мұндағы жарнама мәтінінің сөйлем құрылысы қазақ тіліне жат, орыс тілінен тікелей аударылған немесе сауатсыз құрылған жарнамалар деуге болады. Оның бір себебі қызмет түрлерін ұсынған адамдар өздері қазақ тілінде мәтін құрай алмағандықтан өзгелерге өтініш жасайды, сөйтіп кейде қазақ тілін білмейтін аудармашыларға тап болып, асығыстықпен тексеруден өтпей жарияланып кетеді.

Кейбір жарнамалардың мәтіндері қажетті ақпаратты береді. Мысалы, орыс тілін үйретемін деген мына жарнаманың бірінші сөйлемінде – мақсат, екінші сөйлемінде – қашан, үшінші сөйлемінде – қайда және қандай жерде екені қосымша айшықталып беріле кетеді:

Орысша оқытамын. 1 нші бортевулсинң 2 нші аулында жаз айларында орысша оқушыларды қабұл етемн. Ауыл орны бек көңілді: орман өзен жанында. Поштасы да жақын. Адресім мынау:

Ст. Каменно-Озерная, Оренб. Г. Училию Марианской аульн.школы Малабаеву («Қазақ» газеті, №15, 1913).

20 ғасыр басы қазақ даласында қоғамдық істерді жүргізуге, іскери қағынастар, сот ісі мәселесімен айналысуда оқыған, орыс тілін білетіндердің көмегіне зәру болатын кез еді. Бұл істер бойынша жақын-таныстарға көмек беріп жүрген кісілер енді көптің сұрауына тап болады. Сол себепті «Қазақ» газеті арқылы оқырмандарға өздері туралы құлағдар етуді бастайды.

Кейбір жарнамалар орыс тілінен аударма қалпында жарияланып кеткенмен, қазақ тілінің сөзсаптамы бойынша беріледі. Мысалы, «Г.Загуменва. Русша жазуларды машина мен көшірүб басыб беруді қабұл етеді. Қымбат алмайды. Адресі: Орнборг, Петрнавлофски орам, асдық базарына жақын». Мұндағы қымбат алмау деген тіркес қазіргі уақытта бағасы арзан деген тіркеске ауыстырылған. Шындығында, «қымбат алмайды» деген қолданыс қымбат емес, арзан деген бірліктерге қарағанда қазақы болып табылады.

Денсаулық күтіміне катысты тамыр дәрінің жарнамасы арасына жеті нөмір салып, 3 рет жарияланған, үшеуінде де бір мәтін беріледі, үшеуінде де өнім құны, жеткізу шарты, мекенжайы бірдей:

Шыққан жерінен өзіміз һәр уақыт жаңасын алдырыб тұрғандықтан арзан сатамыз:

Мысқалы 5 тиыннан 70 тиынға қадар һәр түрі бар.

Сомына 25 тиын задатка жіберген кісіге һәр жаққа жіберледі.

Алдырушыларға адресі:

г. Тургай, Абдулкаому Кадышеву («Қазақ» газеті, №7, 1913).

Бұл жарнама тамыр дәрісінің пайдасын білетін, қолданып көрген, қайталап тапсырыс беретін жұртшылыққа арналған. Өйткені мұнда алдыңғы жарнамалардағыдай, сүт мәшиненің, басқа мәшиненің пайдасын алдымен түсіндіріп алған тәсіл қолданылмайды. Бірден өнімнің арзандығын және тез жеткізіп беретіндігін, әрі бағасын жариялап нақты жазады. Бұл да жазу прагматикасын біліп қалған коммуникация ортасы екеіндігін көрсетеді.

Тағы бір ерекшеліп айтатын – газет беті арқылы алғыс айтылатын дәстүр қалыптасқандығы болды. Мысалы, мына жарнамада науқастанып жатқан баланың халін сұраған жалпақ жұртшылыққа көңіл айтылған. Оны «тәңір жарылқасын» деген атаумен айдарлаған:

Тәңірі жарылғасын

Семей Зарешной Ислабодка саудагері Абдулрахман мырза Йусф хажсы ұғлы бұйл науқасқа ілінгеннен бері қарай сыртдан өздері келіб, йаки хат бен көңіл сұраған ағайын-тұусқан, дос-жар жұусқан кісілерге «Қазақ» арқылы

тәңірі жарылғасын айтыб, өзінің ризалғын білдіреді («Қазақ» газеті, №18, 1913).

Мұнда науқасқа іліну, тәңір жарылғасын, дос-жар, жуысқан, ризалығын білдіру деген дәстүрлі сөзсаптамға тән қолданыстар жарнама тілін қазақыландырған.

Жалпы, жарнама мәтінін зерттеушілер жарнама гендерлік жағынан ерекшеленетінін және ерекшеленуі тиіс екенін айтады [280, 59 б.]. Ал «Қазақ» газетіндегі жарнамалар жарнама түзудің алғашқы тәжірибесі болғандықтан, әрі зат сатып алуды ұйғаратын отбасы мүшесі әйел адам емес, үй иесі – отағасы, яғни ер адам болғандықтан жарнама әйел адамға бағытталып жасалмаған.

Зерттеушілер жарнаманың бірнеше типін анықтаған: сауда маркасының жарнамасы, жекелеген сауда жарнамасы, саяси жарнама, анықтамалықтағы жарнама, іскерлік жарнама, институционалды жарнама, әлеуметтік жарнама, интерактивті жарнама, насихаттық жарнама [281, 18 б.]. «Қазақ» газетінің 2013 жылғы нөмірлерінде сауда және әлеуметтік жарнамалар жиі жарияланғаны байқалады және жарнама мәтіндерінің графикалық безендірілуінде де ерекшелік бар, жарнама мәтіндері оқырманның назарын аударатындай айрықшаланып берілген. Негізгі ойды білдіретін сөздер қою әріппен көрсетілген. Бұл ХХ ғасырдың басында жарнаманы безендіру технологиясының айтарлықтай дамығанын көрсетеді. Сонымен қатар, жарнама мәтіндері лингвистикалық тұрғыдан жеңіл әрі қарапайым берілген, бұл кез-келген адамның тез түсінуіне бағытталып жасалған әрекет.

Қорыта айтқанда, «Қазақ» газетіндегі жарнамалар қоғамды ақпараттандыруда тиімді қызмет ете отырып, жарнамаланып отырған нәрсенің өтімді болуына да ықпал ете алған.

10.2 Жарнама мәтіндер ішкорпусын әлеуметтік лингвистика аспектісінде зерттеу

ХХ ғасырдың аяғында компьютерлік технологиялардың және олармен бірге мәтіндік корпустардың дамуы филология ғылымдары саласында нағыз серпіліс жасады. Соңғы жылдары корпустардың қол жетімділігі және олардың құралдарындағы технологиялық прогресс күрт өсті. Қолданбалы лингвистер веб-сайттар, әлеуметтік желідегі блогтар және медиа сайттар арқылы әлемнің түкпір-түкпірінен және әртүрлі тілдерден деректерге кеңірек қол жеткізе алады. Ғалымдар арасында зерттеудің басқа салаларында корпус пен корпус әдістерін қолдануға қызығушылықтың деңгейі артты, атап айтқанда, әлеуметтік лингвистика.

Әлеуметтік лингвистика негізінен әлеуметтік факторлардың тілге қалай әсер ететіндігін зерттейді. Алайда, танымал америкалық ғалым У.Лабов, бұл терминнің «қаншалықты артық» екенін, өйткені тіл мен лингвистика әрқашан әлеуметтік екенін атап өтіп, дәйекті түсініктеме береді [282, 183 б.]. Дегенмен, барлық лингвистер тілдің әлеуметтік аспектілеріне назар аудара бермейді, сондықтан «әлеуметтік лингвистика» термині әлеуметтік контексте тіл үйренуге баса назар аударатын өзара байланысты салалар жиынтығын біл-

діреді деп айтуға болады. Ғалымдар арасында әлеуметтік лингвистика туралы әртүрлі көзқарас бар. Соның бірі, әлеуметтік лингвистика – тілді басты нысана ретінде алып, оны әлеуметтік-қызметтік тұрғыда зерттейтін тіл білімінің бір саласы. Ал екінші көзқарасқа сүйенсек, әлеуметтік лингвистика – лингвистика мен әлеуметтану, тарих, психолінгвистика, этнография т.б. ғылымдардың шекаралық, пәнаралық саласы.

Осы тұрғыдан келгенде, корпус мәселелерін әлеуметтік лингвистиканың тәсілдері мен құралдарын қолдана отырып зерттейтін тіл білімінің ерекше саласы ретінде Корпустық әлеуметтік лингвистиканы атауға болады.

Корпустық лингвистика мен әлеуметтік лингвистиканың екеуі де статистикаға жүгінеді, бірақ олардың зерттеу әдістері әртүрлі. Әлеуметтік лингвистика бір топтан алынған мәліметтерге сүйеніп, белгілі бір нұсқаның орындалу пайызын ескерсе, ал корпустық лингвистика, әдетте, абсолютті жиілікті бағалайды (барлық қолдану жағдайларына немесе барлық мәлімдемелерге қатысты).

Мәселен корпуста ізделетін ақпараттардың (метабелгіленімнің) деректерін адресаттар немесе адресанттар тобы (жынысы, жасы, аймағы, т.б.) немесе мәтін сипаты (ауызша немесе жазбаша) немесе лексемалардың грамматикалық немесе сөзжасамдық сипаттамалары бойынша шектеуге мүмкіндік береді. Бұл мүмкіндік қарастырылып отырған материалдың толықтығын қамтамасыз етеді және сонымен бірге оның параметрлерін тез өзгертуге мүмкіндік береді. Бұл мүмкіндіктер болса, «стандарт шеңберіндегі айырмашылықтарды зерттеу» сияқты корпустық лингвистиканың тағы бір ерекшелігін түсіндіреді. Ал әлеуметтік лингвистикада стандартқа сай келмейтін нұсқаларға жиі қызығушылық танытады [283, 188 б.].

Жарнама қазіргі қоғамның ажырамас бөлігі болып табылады және маңызды экономикалық және әлеуметтік функцияларды орындайды. Алғашқы қазақ жарнамасы ХХ ғасырдың басындағы баспасөз беттерінен бастауын алды. Соңғы жылдардағы практикалық ұсыныстар қазақ жарнамасының теориялық зерттеулерін толықтырды (М.М. Увайсова, А.А. Шортанова, Р.И. Якупова, Ж. Бекболатұлы және т. б.). Алайда, қазіргі жарнама мәтіні кеңестік кезеңдегі жарнама мәтіндерінен ерекшеленеді. Шетелдік жарнама компанияларының тәжірибесі қазіргі қазақстандық жарнаманың қалыптасуына негіз болды. Алайда шетелдік жарнама Қазақстанда әрдайым сәтті бола бермейді. Бұл мәселе қазіргі қазақ қоғамындағы жарнама мәтіндеріне қатысты зерттеулердің қажеттілігін туындатып отыр.

Жарнама мәтіндерін ҚТҰК базасында қарастыру арқылы болашақта қазақ копирайтингінің сәтті болуына жол ашылады. Осы ретте ҚТҰК құрамында әзірленіп жатқан жарнама мәтіндер ішкорпусына тоқталайық. *Жарнама ішкорпусы* – қазіргі қоғамда таратылу орны мен әдісіне қарай бөлінген *көше, көлік, сауда орындары, БАҚ, интернет* жарнамалары мәтіндерінің жиынтығы. ҚТҰК аясында жарнама ішкорпусы Ахмет Байтұрсынұлы атындағы Тіл білімі институтында алғаш рет әзірленіп жатыр. *ҚТҰК-ның құрамында жарнама ішкорпусын құрудың мақсаты* – бизнес коммуникация негізінде өндірістік қатынастарды реттеуші жарнама мәтіндерінің электрондық базасын жасау,

олардың метабелгіленімдерін әзірлеу. Жарнама ішкорпусы *көше, көлік, сауда орындары, БАҚ, интернет* жарнамаларының мәтіндер базасынан, белгіленім, метабелгіленім және іздеу жүйесінен тұрады.

Жарнама мәтіндер ішкорпусын жасау мынадай міндеттерден тұрады.

- 1) жарнама ішкорпусының ауқымды мәтіндік базасын жинақтау;
- 2) жиналған, сұрыпталған жарнама мәтіндер ішкорпусының метабелгіленім, белгіленім, іздеу жүйесін әзірлеу;
- 3) жарнама мәтіндерінің түрлеріне сәйкес модельдер ұсыну;
- 4) ішкорпус базасындағы жарнама мәтіндерін сәтті жарнама, сәтсіз жарнама деп бөліп, қазіргі копирайтерлерге арналған нұсқаулық ұсыну;
- 5) нормативті жарнама маркерлері жүйесі мен әртүрлі нысандар бойынша жарнама модельдері берілген ІТ-әзірleme дайындау.

Жарнама мәтіндер ішкорпусын әзірлеудің бірінші кезеңі жоғарыда келтірілген алғашқы екі міндетті қамтиды. Ең алдымен жарнаманың таратылу орны мен әдісіне қарай *көше жарнамасы, көлік жарнамасы, сауда орындарындағы жарнама, БАҚ-тағы жарнама, интернеттегі жарнама* түрлерін анықтап, тиісті нысандарда берілген жарнамаларды суретке түсіріп, видеосы жазылып, иллюстрациялық материалдар жиналды. Жиналған материалдар сұрыпталып, олардың стенограммасы жазылды. Осылайша жарнама ішкорпусының ауқымды мәтіндік базасы жинақталды.

Көшедегі жарнамада Алматы қаласының 8 ауданы бойынша **1332 жарнама** жиналды. Көшедегі жарнама қалалық инфрақұрылым нысандарында: билборд, баннер, пилон, тролль, брендмауэр, штендер, мультимедиялық экрандар және т.б.; ғимараттарда және сауда нысандарында: ламбрекин, витрина, маңдайшадағы жапсырма, терезедегі жапсырма, қабырғадағы жапсырма түрінде орналасқан.

Көшедегі жарнамаларды жинау барысындағы қиындықтар ретінде көп жерде жарнама қайталанатын, Медеу, Бостандық аудандарында орыс тіліндегі жарнамалар басымырақ екендігі анықталды.

Сауда орындарындағы жарнама (ары қарай – СО-дағы жарнама) Алматы қаласындағы сауда үйлерінен, базарлардан, шағын дүкендерден **1801 жарнама** жиналды. СО-дағы жарнамалар сауда залдарын жарнамалық безендірулерінде; дүкен сөрелерінде (сыртқы және ішкі); сауда залындағы белгілер, маңдайша жазулар және планшеттерде; витрина жапсырмалары; еден жапсырмалары; бликфанг; кассалық аймақтағы жарнама; чекпоинт және т.б. жерлерде орналасқан.

СО-дағы жарнама мәтіндерін жинау барысында Алматы қаласы бойынша 30 ірі сауда, ойын-сауық орындары анықталды. Атап айтқанда: Апорт Молл, Мега Центр (Розыбакиев көш.), Мага центр (Сейфуллин көш.), Алматы Молл, Есентай Молл, Москва, Достық Плаза, Максима, Алма-Ата, Сити Центр, Азия Парк, Март, Домиллион, Ритц Палас, Атакент Молл, Армада, Силквэй Сити, Колибри, Мерей, Прага, Детский Мир, Цум, Гранд Парк, Галлерейя, Восход, Променад, Алтын Тараз, Артриум, АДК, Глобус. Сауда орындарынан кейін базарлар мен кіші сауда орталықтарынан жарнама жиналды.

Көліктегі жарнама көліктерде, вокзалдарда, әуежайда, метрода **700 жарнама** жиналды. Көліктегі жарнамалар көлік құралдарының сыртындағы жа-

зуларда (шатырлар, борт, есіктер, стендтер, әйнектер); көлік құралдарының салондарында орналастырылатын баспа хабарландыруларында; көліктің мониторларында; вокзалдарда, әуежайларда, метрода тауарлары бар витриналарда, мультимедиялық экрандарда; сонымен қатар көлік ішінде орналасқан.

Жарнама ішкорпусының мәтіндік базасын жинақтау мақсатында көліктердегі (жеңіл және жүк автокөліктеріндегі, автобустардағы, метродағы), автовокзалдардағы, ТЖ вокзалдардағы, метро станциялардағы, әуежайдағы қазақша жарнамалар жиналды (суретке және бейне роликке түсірілді). Нәтижесінде 250-ге жуық көліктердегі жарнама, 150 әуежайдағы жарнама, 100 Алматы-2 ТЖ вокзалындағы және 50 Алматы-1 ТЖ вокзалындағы жарнама, 50 автовокзалдардағы жарнама, 110 метро станциялардағы жарнамалар жиналды. Нәтижесінде 550 қазақша жарнама мен 150 орысша жарнама жиналды.

Бұқаралық ақпарат құралдарындағы жарнама (ары қарай – БАҚ-тағы жарнама) бойынша республика бойынша өңірлік теле-радио арналардан **1835 жарнама** жиналды. БАҚ-тағы жарнамалар теледидарда: видео ролик, жүгіртпе жол; радиода: жарнамалық блоктардағы роликтер, жарнамалық хабарлар, ақпараттық және жарнамалық газеттерде: республикалық, жергілікті, кәсіби; журналдарда: арнайы, салалық, экономикалық және т. б. жарнамалар ретінде беріледі.

БАҚ-тағы жарнама мәтіндерін жинақтауды бастаудың алдында Қазақстандағы жарнама нарығына біраз аналитикалық шолу жүргізілді. Шолу нәтижесінде Қазақстандық БАҚ-тағы жарнаманың ақпаратты беру тәсіліне қарай 3 түрлі форматы анықталды:

- аудио жарнама (радио, теледидар);
- визуал жарнама (газет, журнал, интернет-медиа, теледидар);
- аудиовизуал жарнама (теледидар, интернет-медиа).

Осының негізінде жарнама материалдарын жинау жұмыстардың көлемі мен ауқымын білу үшін, БАҚ тізімі жасалды. Аудио жарнама, яғни радио арналар үшін жеке дерекқор; аудиовизуал жарнама, яғни теледидар мен интернет-медиа үшін жеке дерекқор; визуал жарнама, яғни газет, журнал, интернет-медиа үшін жеке дерекқор жасалды. Нәтижесінде 35 радио арна, оларды басқаратын 17-ге тарта компания анықталды. Қазақстан Республикасы аумағында жұмыс жасайтын 35-тен аса телеарна табылды. Газет, журнал, жаңалықтар порталдарына қатысты ақпарат жинақтап үлгермеді. Себебі баспа БАҚ саны, олардың түрлері және жарнама саясаттары туралы ақпаратты ашық қорлардан табу қиындық тудырып жатыр. Жинақталған дерекқор негізінде Қазақстандағы радио арналарды басқаратын 16 компанияға хат әзірленді.

Интернеттегі жарнама бойынша Instagram, Facebook, YouTube және басқа сайттардан **400 жарнама** жиналды. Интернеттегі жарнамалар сториз, пост, таплинк, бот, ролик, дайрект арқылы беріледі.



Сурет 26. Жарнаманың мәтіндік базасы бойынша жалпы статистикасы

Қазіргі таңда әрбір жарнама мәтініне метабелгіленім қойылып жатыр. Жарнаманың мәтінінің метабелгіленімі 13 параметрден тұрады. Олар:

– **жарнаманың материалы.** Бұл жерде көше, көлік, сауда орындарындағы жарнамалардың суреті беріледі.

– **жарнама мәтінінде** суреттегі жарнаманың толық мәтіні қойылады.

– **жарнаманың түрі** деген параметрде жарнаманың таратылу орны мен әдісіне қарай бөлінген түрлерінің бірін жазады, яғни берілген жарнамаға қарай *көше, көлік, сауда орындары, БАҚ, интернет* жарнамаларының бірі жазылады.

– **жарнама тақырыбы** өз ішінде макротоп пен микротоп ретінде бөлінді. Макротопқа жарнамаланған заттың (өнімнің, тауардың) жалпылама атауы, микротопқа жарнамаланған заттың (өнімнің, тауардың) жеке атауы көрсетіледі. Мысалы, әйгілі Соса Сола туралы жарнама мәтінінің метабелгіленімінде макротоп – «Сусын» болады да, микротопқа тауар маркасы, яғни «Coca Cola» жазылады.

– **жарнаманың формасы** параметрінде жарнаманың берілуіне қарай «визуал», «аудио», «аудио-визуал» формаларының тиісті түрі жазылады.

– **жарнаманың типі** әр жарнама түрінің орналасуына қарай ерекшеленеді. Мәселен, көше жарнамасы билборд, баннер, пилон, сауда орындарындағы жарнама бликфанг, чекпоинт сияқты т.б. құрылғыларда беріледі. Міне, осылар жарнаманың типі ретінде белгіленеді.

– **жарнаманың жылы** міндетті түрде көрсетіледі.

– **жарнама беруші** ретінде жеке адам / мемлекеттік мекеме / қоғамдық ұйым / шағын бизнес / орта бизнес / ірі бизнестердің бірі көрсетіледі.

– **берілу аймағына** жарнаманың жарияланған мекенжайы жазылады.

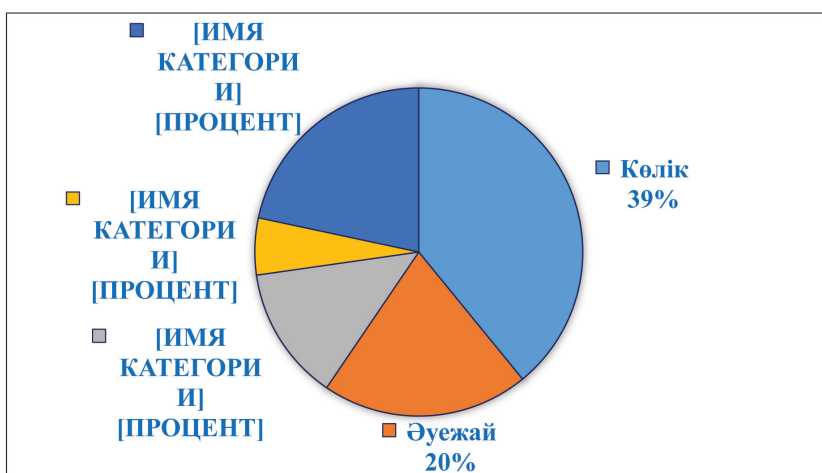
– *кері байланыс* жарнама мәтінінің метабелгіленімі үшін өте маңызды. Мұнда жарнама жариялаушы тараптың байланыс түрлері беріледі.

– *мәтінді белгілеуші* параметріне жарнама мәтінінің метабелгіленімін жасап отырған тұлғаның аты-жөні жазылады.

– *корпусқа енгізілу уақыты* туралы мәлімет беріледі.

Жарнама мәтіндерінің метабелгіленіміне жазылатын бұл параметрлер корпусық әлеуметтік лингвистиканың маңызды бір нысаны болып табылады. Сонымен қатар зерттеудің шынайылығына, нақты әрі дәлелді болуына мүмкіндік береді.

Жоғарыда айтып өткен жарнама түрлерінің ішіндегі **көлік жарнамасы** халыққа визуалды жарнама ұсына отырып, ерекше әлеуметтік мәнге ие болады. Көліктегі жарнамалар екі түрлі нысанда жарияланады. Бірі көліктегі жарнама болса, екіншісі – қоғамдық көліктердің және жолаушы тасымалдайтын транспорттардың (автобус, пойыз, ұшақ, метро) орналасқан нысандарындағы жарнамалар. Мұндай нысандарға метро станциялары, автобөкеттер, әуежай және теміржол вокзалдары жатады. Бұл жерлерден сурет және бейне (видео) материалдар түрінде кездескен жарнамалар жиналды. Әр материал белгілі бір жарнама беруші жекеменшіктің немесе мемлекеттік мекеменің жарнамасы болып табылады.

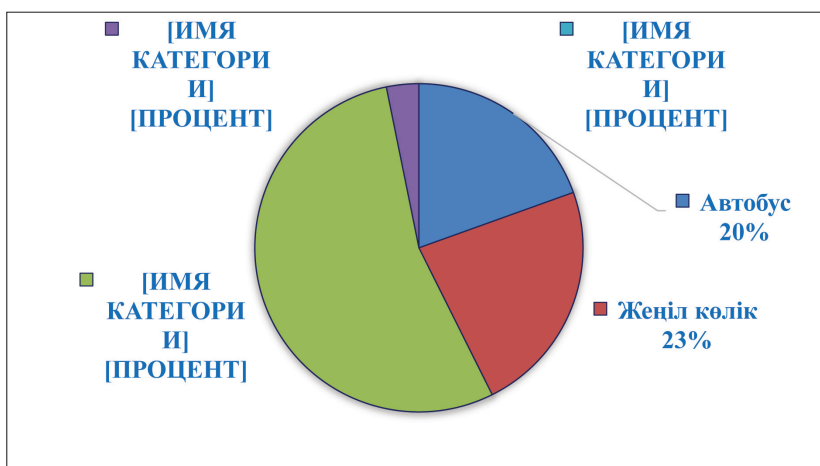


Сурет 27. Көлік жарнамасының жалпы статистикалық көрсеткіші

Көліктегі жарнамалар Алматы қаласындағы көліктерден, көлікке жарнама жапсыратын компаниялардың интернеттегі сайттарынан, әлеуметтік желідегі парақшаларынан алынды. Жарнаманы көлікке жапсыруды «брендинг» деп атайды және ондай жекеменшік компаниялардың қатарында Алматы қаласындағы «Рекламный советник», Gmask, AVP pro, Live adver т.б. бар. Ол жарнамалардың интернетке жарияланған уақыттары және орналасқан қалалары әртүрлі. Ал жолаушы тасымалдайтын көліктерден 2023 жылдың ақпан-наурыз айлары аралығында ілінген жарнамалар жиналды. Сурет және бейне материалдар түріндегі жарнамалардың сапалық тұрғыдан сұрыпталып,

сандық деректері анықталды. Келтірілген статистика суреттердегі қазақша және орысша жарнамаларға негізделеді.

Көліктің өзіндегі жарнамалар жапсырылған жеріне қарай **көлік ішіндегі** және **көлік сыртындағы жарнама** деп екіге бөлуге болады. Көлік ішіндегі жарнамалар көп жағдайда қоғамдық көліктерде болады. Ондай жарнамалар бүгінде өте аз кездеседі. Алматы қаласы бойынша қоғамдық көлікте кездесе бермейді. Мұндай жарнамаларды арнайы интернеттен іздеу барысында, республика бойынша тек 8 жарнаманың суреті табылды. Мысалы: «*Ақша қажет пе? Онлайн шағын несие ал*», МҚҰ «*Эверест финанстың*» Төмен % мәлішерлемелер дейін 1 000 000 тг», «*FlyArystan компаниясының Ақтөбе-Атырау 4999 т* бастан*» т.б. Бұл мысалдарда тек басты үлкен қаріптермен жазылған жарнамалар ғана көрсетілген.



Сурет 28. Көлік ішіндегі жарнаманың статистикалық көрсеткіші

Ал көліктің сыртына жапсырылған жарнамалар сан жағынан әлдеқайда көп. Оларды көліктің түрлеріне қарай **автобустағы жарнама, жеңіл көліктегі жарнама, жүк көлігіндегі жарнама** деп қарастырдық. Бұл көліктерде мынадай жарнамалар кездеседі: FLASH ENERGY-дің «*Қуатым бар!*», YET.KZ-тің «*YET! ЕТКЕ Еттен жасалған шұжықтар*», Канц TREND-тің «*Кеңсе тауарлар әлемі*» т.б. Жарнамалардың ішінде латын әліпбиі негізінде жазылған жарнамалар да бар. Мысалы, Arb.kz-тің «*Jol talğamaityn kólikter men pikaptar tiýningi*», Jpi.kz-тің «*Japondyq jaialyqtar Momi*», Megawatt-тың «*Akkýmýliator ortalıǵı*» және т.б. Мұндай жарнамалардың жалпы саны – 15. Автобустағы жарнамалардың саны – 49, жеңіл көліктегі жарнама саны – 58, жүк көлігіндегі жарнамалар саны – 136. Көлік сыртындағы жарнамалардың жалпы саны – 243 болды. Бұл көрсеткіштен тауарларды тасымалдайтын жүк көлігінде көп кездесетінін көруімізге болады. Біздің пікірімізше оның себебі, жүк көлігі тауар тасымалдағандықтан, қаланың түкпір-түкпірінде көп жүреді және жүк көлігінде жеңіл көлікке қарағанда жарнама орны үлкен. Ал Алматы қалалық автобустарға бұрынғыдай жарнама жапсыруға рұқсат берілмейді. Тек авто-

бустарды шығаруды қаржыландыратын «Өнеркәсіпті дамыту қоры» АҚ-тың белгісі мен интернеттегі сайтының сілтемесі тұрады.

Әуежайда жарнамалар сырттағы және іштегі банерлерде, банкоматтарда орналасқан. Мысалы, сырттағы жарнамада ҚР ПМ-нің «*Жылдамдықты азайт – өміріңді сақта*», Рухани жаңғыру бағдарламасының «*«Біз ХХІ ғасырдың жаһандық картасында ешкімге ұқсамайтын, дербес орны бар ұлт боламыз десек, әлем бізді мәдени жетістіктерімізбен тануы керек» Н.Назарбаев*» сынды әлеуметтік жарнамалар, сондай-ақ Olimpbet, шетелдік әншілердің жарнамалары, «Моншалардың жоғарғы лигасы» сынды т.б. жарнамалар ұшырасады. Ал әуежай ішінде Халық банкінің, Costa coffee, Visa, әуе компанияларының жарнамалары, Нұрбанк, Жусан банк, Центркредит банктерінің банкоматтарындағы экономикалық жарнамалар, Covid-19 вирусынан қорғануға байланысты әлеуметтік жарнамалар және әуежайдың ресми таксиінің жарнамалары да бар.

Метро станцияларындағы жарнамалар станцияға кіреберістегі банерлер мен сауда бутиктерінде және іштегі банерлерге ілінген. Әртүрлі Технодом, Burger king, Masscoffee сынды көптеген ірі компаниялардың жарнамалары кездеседі. Әлеуметтік жарнамалар өте аз. Станцияларда және метроның ішінде LED-экрандарда көп жарнамалар көрсетіліп тұрады. Led-экрандардағы жарнамалардың көп бөлігі метро станциялары мен әуежайда кездеседі. Жарнамалардың сандық көрсеткіші келесідей: Алматы-1 теміржол вокзалы – 1, әуежай – 9, метро станциялар мен метрода – 17. Барлығы – 27. Бұл сандық көрсеткіштер жарнамалардың санын емес, түсірілген бейне роликтердің (ішінде бірнеше жарнамалар бар) санын көрсетеді. Бейнероликтегі жарнамаларға Cosmart, Altel, LG, MYD production, т.б. компаниялар мен концертке шақыру билеттері көрсетіліп тұрады.

Вокзалдар мен автобекеттерде сауда бутиктерінің банерлерінде, ойын автоматтарында, кофе дайындайтын, сусын, шоколад сататын және жолақысын тұрақты төлейтін құрылғыларда кездеседі. Автобекеттерде ешқандай LED-экрандар кездеспейді. Онда жарнамалардың көп бөлігін сауда бутиктерінің жарнамалары алады. Арасында садақа берудің және президенттікке сайлау сияқты әлеуметтік жарнамалар кездеседі. Вокзалдарда әлеуметтік жарнамалар өте көп. Мысалы, Алматы көлік прокуратурасының, ҚР ПМ-нің, президенттікке және парламентке сайлауға үгіттеудің, өрттен және вирустан сақтанудың т.б. жарнамалар. Ойын автоматтарындағы, Оңай жүру ақысы, тұрақ үшін төлеу жарнамаларда нұсқаулық жазылған. Ал сусын мен тамақтанатын заттарды, кофені сататын құрылғыларда нұсқаулықпен қатар жарнамалар да кездеседі. Мысалы, Пекарня мельницаның «*Нан өнімдері*», KitKat-тың «*Үзіліс бар, KitKat бар*».

Жарнамалар жолаушы тасымалдайтын орындарда жекеменшік компанияларының жарнамасынан бөлек әлеуметтік жарнамалар қоғамдық жаңалықтарға байланысты өзгеріп отырады. Мысалы, 2023 жылдың 19 наурызында сайлауға байланысты партиялар мен депутаттар үгіт-насихат жүргізіп, жарнамаларын ілген болатын. Мысалы: «*Жер тек жерде еңбек еткенге берілсін! Қолдаймын! Қазақстан Халық партиясы*», «*Атапат Атапат-қа дауыс бер. 19 наурыз*», «*Батыл қадам батыл қоғам заманы Алматылық Ерлан Стамбеков*» т.б.

Сонымен қатар көлік жарнамасы бойынша тек орыс тіліндегі жарнамалар өте аз кездесті (20-кесте). Себебі көп жағдайда жарнамалар қазақша-орысша бірге беріледі.

Кесте 20

Көлік нысандарында орыс тіліндегі жарнамалардың статистикалық көрсеткіші

Орысша жарнама түрлері	Саны
Көлік	8
Әуежай	8
Метро	14
Сайран автобекеті	12
Алматы-1 вокзалы	14
Барлығы	56

Дегенмен тек қана орыс тілінде берілген жарнамалар Қазақстан Республикасы Тіл туралы заңның 21-бабына қайшы келеді. Онда былай делінген: «Егер Қазақстан Республикасының заңнамалық актілерінде өзгеше көзделмесе, хабарландырулар, жарнама, прейскуранттар, баға көрсеткіштері, ас мәзірлері, нұсқағыштар және басқа да көрнекі ақпарат мемлекеттік тілде, қажет болған кезде орыс тілінде және (немесе) басқа тілдерде де орналастырылады» [284]. Демек жарнамалар тек қазақ тілінде немесе қазақ тілімен бірге орыс немесе өзге тілдерде орналастырылуы керек.

Жарнамалар мазмұны жағынан әр түрлі болып келеді. Тұтынушыға сұрақ қою арқылы басталатын жарнама моделі едәуір. Мысалы: *Ақша қажет пе? Оңайырағын қалайсыз ба? Егер тегін 30 ГБ болса, трафикке ақы төлеп керегі не? Егер тегін 30 ГБ болса, трафикке ақы төлеп керегі не?* т.б. Сонымен қатар, эмоцияналды-экспрессивті реңктегі көтеріңкі көңіл кұйді білдіретін леп белгісімен жазылған жарнамалар өте көп. Мысалы: МҚҰ «Эверест финанстың» *«Тез және ыңғайлы!»*, Pin-up.kz *«Ойнап өмір сүр!»*, Eurasian food *«Әрбір үйде күн сайын!»*, Sibylla *«Махаббатпен Швециядан!»*, Империя сыра *«Әр тілімдегі рахат!»* т.б.

Қорыта айтқанда, жарнама мәтіндерінің түрлерін жүйелеу және статистикалық көрсеткішін анықтау, жарнама ішкорпусын құрып, онда жарнама мәтіндерін жинау, қазіргі копирайтерлерге арналған нұсқаулық ұсыну бүгінгі күні өте маңызды жұмыс болып отыр. Себебі жарнамалар халыққа ұсынылғаннан кейін, оның сапасы жоғары, тілі көркем әрі халыққа түсінікті болуы тиіс. Біз осы зерттеу жұмысы арқылы жарнама түрлерін және олардың ерекшеліктерін, статистикалық көрсеткішін анықтадық. Жарнама ішкорпусын одан әрі кеңейту арқылы, жарнама мәтіндерін тілдік тұрғыдан зерттеп, сәтті және сәтсіз жарнамаларды анықтау алдағы зерттеу жұмыстардың өзекті мәселесі болады.

10.3 Қазақ тіліндегі жарнама мәтіндерінің тілдік ерекшеліктері

Бүгінгі таңда тілдің коммуникативтік әлеуетін танытатын ақпарат ағыны арналарының бірі – жарнама саласы. Өндірушілер мен тұтынушылар қатынасын сипаттайтын нарықтық экономика жағдайында жарнама саласының қоғамға ықпалы күннен күнге артуда. Жарнама – тауар немесе қызмет сатылымына әсер етуші, сатылым көлемін арттырушы негізгі құрал болып қана қоймай, қоғам психологиясын терең түсінуді қажет ететін күрделі ұғымға айналып отыр. Осыған орай, жарнама қызметі, құрылысы, әсіресе жарнама мәтіні түрлі ғылым салаларының аясында кеңінен қарастыруды, зерделеуді қажет етеді. Қазақ тілінің түсіндірме сөздігінде жарнама сөзіне келесідей анықтама берілген. «**ЖАРНАМА** [ир.] з а т. <франц. reclame> **1.** Тұтынушылардың назарына іліктіру немесе қандай да бір сауалнама алу мақсатында тауар, өнер туындылары және қызмет түрлерінің қасиеттері туралы кең таратылып жария етілетін хабарлама. **2.** Тұтынушылардың назарына іліктіру мақсатында тауар, өнер туындылары және қызмет түрлерінің қасиеттері туралы хабарламаны таратудың қағаз, радио хабарлама, бейнешығыршық, жазба мәтін т.б. түріндегі құралы. **3.** Мандайшаларда берілетін ұйым не мекеме туралы ақпарат. **4.** Афиша спектакль, концерт, жиналыс, дәріс, спорттық жарыстар, ойын-сауықтар туралы қағаз хабарлама. **5.** Танымалдығын арттыру, жұрт назарына ілігу мақсатында біреу немесе бір нәрсе туралы таратылған мәліметтер» [285, 213]. Берілген анықтамада жарнама ұғымы терминдік ерекшелігіне сай барынша кеңінен қарастырылған. Жарнаманың негізгі қызметі ақпарат беру және әсер ету екенін ескерсек, жарнама мәтіні жоғарыдағы қызметті іске асыратын негізгі құрал болып табылады. Жарнама мәтіні ретінде көбіне жарнаманың вербалды мазмұны түсініледі [286, 5]. Жалпы тіл білімінде жарнама мәтіндері бұқаралық ақпарат құралдары арқылы кең таралатын, тіл қолданушының сөз саптауына, сөздік қорына, сөйлеу мәдениетіне бірден-бір әсер етуші коммуникация түрі ретінде бағаланады. Олай болса, жарнама мәтіндерін зерделеуде бірінші кезекте тіл мәдениеті, ұлттық нормалардың сақталуы, стилистикалық ерекшеліктері, мәтіннің прагматикалық қуаты сынды бірқатар мәселелер өзектенеді.

Жарнама мәтінін тілдік тұрғыдан талдамас бұрын, сапалы әрі өтімді жарнама мәтініне зерттеушілер тарапынан қойылатын негізгі талаптарды айқындау маңызды. Жарнама тілі мен жарнама стилистикасын зерттеуші Д.Э. Розенталь [287, 5] мен Н.Н. Кохтевтің [288, 3], В.В. Ученова [289, 298], Л.Г. Фещенконың [290] еңбектерінде жарнама мәтінінің әсер ету қызметіне баса назар аударылады. Жарнама – аларманды тауар сатып алуға және тапсырыс беруге итермелейтін қуатты құрал. Жарнаманың өтімді болуы үшін тілдің барлық мүмкіндіктерін кеңінен қолдану қажеттілігі баяндалады. Жарнама мәтінінде адресатқа бағытталған коммуникацияның бірінші орынға шығуы байқалады. Олай болса, жарнама мәтініне қойылатын талаптардың қатарында адресаттың назарын аудару, әсер ету мақсатында тілдің көрнекі құралдарын, бейнелі орамдар, экспрессивті-эмоционалды лексикалық қорын және т.б. таңдау, іріктеу және ұтымды қолдану мәселесінің де маңыздылығы айқындалады.

Н.Г. Иншакова жарнама мәтінін құрастырушы мамандардың назарына жарнама қызметін терең түсіну және жарнамаланатын тауар немесе қызмет жайлы барынша мол ақпарат жинау қажеттілігін ескертеді. Сондай-ақ, сапалы әрі өтімді жарнама мәтініне қойылатын келесідей талаптарды көрсетеді.

- Біріншіден, жарнама мәтінінің заңнамалық құқықтық талаптарға сай болуы («Жарнама туралы» заңның негізгі талаптарына қайшы келмеуі);
- Екіншіден, жарнама мәтінінің бірегей сауда ұсынысы талаптарына сәйкестігі (коммерциялық орындығы, тұтынушыға әсер ете алуы, тауарды немесе қызметті алуға қызығушылығын оята алуы);
- Үшіншіден, жарнама мәтінінің орфоэпиялық, орфографиялық, сөз мәдениеті нормаларына сәйкестігі (сөз дұрыстығы, сөз тазалығы, сөз дәлдігі, сөз әсерлілігі т.б.) [291, 32].

Жарнамалық мәтіндерді жасау және түзету кезінде жоғарыдағы ережелерді сақтамау, сауатсыз әрі әсерлілігі төмен жарнама мәтіндерін тудырып қана қоймай, сонымен бірге қоғамдық өмірдің этикалық және мәдени нормаларына қайшы келетін ақпараттың жаппай тарауына алып келуі мүмкін. Сол себепті, нарықтық бәсекелестік жағдайында жарнаманың тиімділігі үшін жарнама мәтініне қойылатын талаптардың қатандануы да заңды.

Ендігі кезекте қазіргі қазақтілді қоғамның ақпарат кеңістігінде жарнама мәтіндері көрсетілген талаптарға сай ма? Қазақтілді тұтынушыларға өзіне қажетті тауар жайында сапалы жарнама мәтіндері қолжетімді ме? деген орынды сұрақтар туындайды. Қазақстан Республикасының «Жарнама туралы» заңының 2-тарауының 6-бабында жарнамаға қойылатын жалпы талаптардың ішінде «Қазақстан Республикасының аумағындағы жарнама (мерзімді баспасөз басылымдарын, интернет-ресурстарды, ақпараттық агенттіктерді қоспағанда) қазақ тілінде, ал жарнама берушінің қалауы бойынша орыс тілінде және (немесе) басқа тілдерде де таратылады» [292] деп көрсетілген. Олай болса, кез келген жарнама беруші үшін (мерзімді баспасөз басылымдарын, интернет-ресурстарды, ақпараттық агенттіктерді қоспағанда) тауарларын немесе ақыға ұсынатын қызметтерін мемлекеттік тілде жарнамалау міндетті болып табылады. Осы талапқа сай әлемдік алпауыт өндіруші компаниялар тауарларының жарнамасын қазақ тіліне аударып жарнамалап жүр. Қазақ тіліндегі жарнама мәтіндерінің көптеген тілдік мәселелері де дәл осы аударманың дұрыс болмауынан, синтаксистік нормаға сай құрылмауына байланысты туындап отыр.

Өзге тілді жарнама мәтіндерін қазақ тіліне аудару барысында қазақы танымға сай мағына бере алмайтын, тілдік құрылысы қазақы сөзсаптамға жатпайтын /сай келмейтін жасанды калька қолданыстар көптеп қолданылады. Жарнама саласындағы мемлекеттік тілдің қолданылуына қатысты аталмыш мәселелер бұған дейін де тілші ғалымдар тарапынан ұдайы айтылып, тіпті бірнеше зерттеу жұмыстарының нысанына айналып үлгерді. Мемлекеттік тілдің жарнама саласына қатысты қолданылу аясын кеңейту қажеттігін және аударма тілін сапалы ету жөнінде: «Мемлекеттік тілде берілген қай жарнаманы алсақ та, талғамсыз, қазақ тілінің әдеби нормаларына қайшы келетін, не болмаса орыс тілінен қалай болса солай аударған жарнамаларға кез боламыз. Қазақ тілінің мемлекеттік тіл дәрежесін толығымен атқаруы үшін оның жар-

нама саласында кенінен қолданылуын, өрісін кеңейту өзекті мәселе... Нарық қыспағына түскен қазақ тілінің мемлекеттік тіл ретінде толыққанды жарнама тіліне айналдырудың тетігін (механизмін) жасауға шамасы келгенше күллі тілші қауым атсалысуы тиіс» [293, 13] – деген пікірлер кездестіреміз.

Тұтынушылық қажеттілікті қанағаттандыру үшін күнделікті түрлі елдер арасында сауда байланысы өсуде. Жарнама тілінің ықпалы әлдеқашан артып, бүгінде жарнаманы бір тілден екінші тілге аудару арқылы өзге тілді мәдениет пен әр түрлі құндылықтарды, әдет-ғұрыптарды еркімізден тыс менгеріп жүргеніміз айқын. Олай болса, жарнама аудармасын мәдениетаралық коммуникация құралы ретінде тануымыз да орынды. Бұл жарнама мәтіні аудармасының сапалы әрі танымды болуын одан әрі өзектендіреді. Жарнама аудармасының сапалы болуы жарнаманың прагматикалық қызметіне де елеулі әсер етеді. Ғ.Ә. Солтанбекова қазақ жарнамасының прагматикалық қуаты жөнінде келесідей пікір білдіреді. «Басқа жазба мәтіндерге қарағанда, жарнама мәтінінде жаңа қолданыстарды қолданудың жауаптылығы жоғары. Жарнама мәтініндегі жаңа қолданыстың тосын екендігі прагматикалық қызметті төмендетеді. Сонымен қатар кейбір жаңа қолданыстар өзінің семантикасы жағынан орыс тіліндегі сөздердің мағынасына сәйкес келмейді. Бұл да жарнаманың тілдік сипатындағы прагматикалық қызметіне әсер етеді [294, 99]. Олай болса, жарнаманы бұқаралық коммуникацияның бір түрі ретінде, қазақ тілінің мемлекеттік мәртебесінің артуына, бекуіне бірден-бір әсер етуші сала екенін ескере отырып, жарнама тілінің қазақылануы, жарнама аудармаларының сапалы әрі өтімді болуы – қазақ тілді қоғам үшін маңызды мәселе деп есептейміз. Осы тұрғыда қазақ тілді жарнама мәтінінің сауаттылығын, әдеби тілдің нормаларының сақталуын, аударма мәтінінің сапасын қадағалап отырудың маңызы зор. Бұл мәселе өз шешімін қазақ тілін цифрландыру жобасы толықтай іске асқанда таппақ. ҚТҰК қазақ тілді жарнама мәтіндерімен толығырақ түсуі қажет. Жарнама корпусы – қазақ тілді жарнама мәтіндерін жинақтау арқылы, келешекте жарнама мәтіндерін сапалы әрі өтімді етуге көмекші таптырмас құрал болмақ. 2023 жылдан бастап А.Байтұрсынұлы атындағы Тіл білімі институтының бастама-сымен жүзеге асып жатқан «Қазақ тілінің ұлттық корпусына» жобасына жарнама корпусы материалдары да жинақтала бастады.

Жарнама корпусының базасын жинақтау барысында басқа кәсіп түрлеріне қарағанда ірі бизнес жарнама мәтіндерінің анағұрлым сауатты құрылатынын аңғардық. Өйткені жарнама маркетингін ірі бизнес қана жасай алады, сондықтан да оны сауатты етіп жасауға көбірек мән береді. Алайда сәтсіз жарнамалар да жоқ емес.

Бірнеше брендтің басын құрайтын алпауыт әлемдік **Pepsi** компаниясының «**Празднуй по вкусу!**» деген жарнамасы билбордта «**Дәмине қарай тойла!**» деп беріліпті. «Той» түсінігінің қазақтардың өмір салтында ерекше орын алатынын ескерсек, жарнама берушінің «той тойлау» тіркесін алуы өте сәтті шыққан. Бірақ бұл аударманы «Дәмін татып, той тойла» деп берсе болатындай. Өйткені «дәм тату» да халқымызға ғана тән дәмге қатысты жоралғы болып саналады.

Toyota Center Almaty ірі бизнес компаниясының «**Большой выбор автомобилей с пробегом**» дегеннің қазақ тіліндегі аудармасын «**Жүрілген авто-**

көліктердің үлкен таңдауы» деп беріпті. Мұндағы «үлкен таңдау» тіркесін алып қарайтын болсақ, сөздердің арасында заттық-логикалық байланыстың жоқтығы байқалады. Өйткені таңдаудың үлкен-кіші көлемде болмайтыны белгілі. Ал «кең таңдау» деп берсек, сөздердің мағыналық жағынан сәйкестігі, яғни зат есім мен сын есімнің үйлесімділігі негізінде тіркес жасалды деп айтуға әбден болады.

Қазақстандағы ұялы байланыс операторларының жарнамалары көптің назарын аудартпай қоймайды. Өйткені олардың қызметін халықтың барлығы дерлік тұтынады. Мәселен, **Tele 2 ұялы байланыс операторының** билбордтағы мына бір жарнамасына назар аударайық: *«Дольше с нами – больше ГБ»* деген сөйлемді *«Қанша ұзақ болсаң – сонша көп ГБ береміз»* деп аударыпты. Бір қарағаннан-ақ калька аударма екені көрініп тұр. «Біздің қызметті көбірек тұтынсаңыз, сізге көбірек ГБ береміз» деген ойды білдіргісі келгенге ұқсайды.

Кең танымал **Dizy club** компаниясының *«Балуем своих!»* деген жарнамасының қазақша баламасы *«Өзіміз зой, еркелейік!»* деп беріліпті. Т.Ф. Ефремованың «Современный толковый словарь русского языка» сөздігінде бұл сөздің мағынасын «близкие», «родные», «родственники» деген сөздермен аша түскен. Демек, *«Өз жақындарымызды еркелетеміз!»* деу анағұрлым сәтті балама болар еді.

Тағам өндісінде кең танымал **«Доширак»** компаниясының мына жарнамасына назар аударайық: *«Сенің үздік таңдауың корей халқының шынайы дәстүрімен және корейлік технологиямен жасалған. Премиум-класс раменының дәмін татып көріңіз»*. Тек тыныс белгісін ұмыт қалдырғаны болмаса, жарнама мәтінінің сапасы жаман емес. Алайда сөйлемнің ұзақтығы оны қабылдауға кедергі келтіретін сияқты. Бұған қоса «шынайы дәстүр» тіркесінен сөздің лексикалық тіркесімділігінің бұзылуы байқалады. Оны «ұлттық дәстүр» деген тіркеспен ауыстыруды ұсынар едік. Өйткені бұл жерде әңгіме корей халқына қатысты болып тұр.

Енді өзіміз күнделікті тапсырыс беретін **Яндекс@еда** жылдам тапсырыс беру қызметінің жарнамасына назар аударайық: *«Тағы да көп жаңа мейрамханалар»*. Жарнама мәтіні, әдетте, атаулы сөйлем формасында, толымсыз формада келеді десек те, бұл жерде грамматикалық жағынан сөйлем құрылымының дұрыс еместігін байқаймыз. Оның орнына «Жаңа мейрамханалар» деп берсе болатын сияқты, өйткені әрі қарай сол жаңа мейрамханалардың ас мәзірімен танысуға мүмкіндік алады.

Ал түркиялық **Ulker Laviva** деген атпен танымал азық-түлік компаниясы *«Насладись изысканным вкусом»* дегенді *«Таңдаулы дәмнен ауыз тиіңіз»* деп беріпті. Орысшадағы «изысканный вкус» «сирек кездесетін», «жасалуы қиын», «тек жоғары тап өкілдеріне қолжетімді» деген мағынаны білдіреді. Бұл батыс мәдениетіне қатысты бірліктерді қазақ тілінің лексикалық қорына ендіру процесі болуы да мүмкін деп топшылаймыз. Жалпы тағамға қатысты бір мәселе: қазақ халқында технологиясы жағынан жасалуы қиын тағамдар жоқтың қасы, ал Еуропада бұл сәл басқаша. Мысалы, вино жасау технологиясы, атақты голландиядық ірімшіктердің жасалу технологиясы да көп уақытты, көп қол еңбегін және көп адамдардың күшін қажет етеді. «Изысканный» сөзінің

шығу төркіні осыдан туындауы да әбден мүмкін. Ал қазақ халқында «таңсық ас» деген қолданыс бар. Оның мағынасы сәл басқа. Белгілі бір өнімнің адамдарына басқа елдердің тағамдарының бейтаныс болуы. Бұл жерде «таңдаулы» ұғымы қазақтың танымына сәйкес келмейтіндей көрінеді. Өйткені қазақ ежелден айналасына қамқор, қолындағы барымен бөлісуге асығып тұратын халық. Жанындағылардың жайын ойламай, өзінің «таңдаулы» ас ішіп отырмағаны анық.

Аты аңызға айналған **Tissot** сағат брендінің «Уақыт тарту етеміз» деген жарнамасына назар аударып көрейік. Бұл мәтін қазақтілді тұтынушыны еш қызықтыра алмайды әрі мұнда ешқандай логика байқалмайды. Өйткені қазақ халқында «сый тарту ету», «ән тарту ету» деген тіркестер бар, бірақ ешқашан уақытты тарту ету деп қолданбаған. Оның орнына «Уақыт алтыннан да қымбат», «Кешегі күннен алыс жоқ, Ертеңгі күннен жақын жоқ» немесе «Уақыт – көшкен керуен» деп келетін қанатты сөздер мен мақал-мәтелдерді қолданса, қазақтың лингвомәдени бірліктері қазақтың танымын айқындар еді.

Қазақстанның ірі сауда компаниясы, электроника мен тұрмыстық техниканы сатудағы көшбасшы **Sulpak** компаниясының жарнама мәтініне назар аударып көрейік: **Выгода в каждой покупке! / Пайда әр сатып алуда!** Орыс тілінен тікелей аударса салғаны байқалып тұр. Қазақ тілді тұтынушы үшін қазақы танымға сай қолданыстарды жұмсау әлдеқайда өтімді болады деп есептейміз.

Centercredit банкі мына бір жарнама мәтінін сәтті беріпті: **Лучший курс в приложении BCC.KZ» / «Ең үздік бағам BCC.KZ қосымшасында.**

«**In Clinic**» деген атаумен қызмет көрсететін тіс емдеу кликасы мынадай мәтінмен тұтынушыны өзіне тартуды мақсат қылыпты: **Улётная улыбка! / Ұшыратын күлкі!** Әрине, калька аударма. Қазақ әдеби тілінің сөздігінде күлкінің бірнеше түрі берілген: наз күлкі, шат күлкі, жайраң күлкі, аппақ қардай ақ күлкі, т.б. Демек, бұл сөздің мағынасын дәл беретін балама да қазақ тілінде өте көп екенін аңғарып отырмыз.

Жарнама – қазіргі таңда динамикалық тұрғыдан ең жылдам дамып жатқан салалардың бірі. Мәтіні, ақпарат беру формасы, ойды жеткізу тәсілдері, рәсімдеу форматтары тұрғысынан әралуандыққа ие құралдардың жиынтығы жарнама саласында бар. Соның арқасында жарнама күнделікті тіршілігіміздің түрлі аспектілеріне еніп кетті. Жарнама мәтіні – өз бойында белгілі бір уақыт пен кеңістік, әлеуметтік формация, қоғамның құрылысы, қоғамның стратификациясы, даму деңгейі, қоғамда жүріп жатқан әлеуметтік үдерістер, жергілікті құбылыстар, жаһандық үдерістер, мәдени код, қоғамда қалыптасқан идеологиялық көзқарастар, діни таңдау және өзге көптеген ақпаратты қамтитын күрделі контент түрі. Сол себепті жарнама мәтіндеріне қатысты екі түрлі тенденцияны байқаймыз. Бірінші тенденция – жарнама ақпараттық кеңістігіміздегі ең көп тараған контент түрі. Жарнама адам баласы туғаннан бастап, өмір бойы айналасында болатын ақпарат түрі. Бұл ақпарат ол адамның таңдауына әсер етіп, белгілі дәрежеде тіршілігіне ықпалын тигізеді. Оның үстіне әлем бойынша үдеуі артып, жаңаша күшке ие болған жаһандану мен цифрландыру үдерістері жарнаманың цифрлық кеңістікке агрессивті түрде енуіне себеп болып жатыр. Бұл тұрғыдан жарнаманың алдағы уақытта адам тіршілігінің

ажырамас бөлігіне айналатынын көрсетіп отыр. Екінші тенденция – жарнама актуалдылығы мерзімінің қысқаруы. Ақпаратты жеткізу технологияларының жылдамдауы жарнама мәтіндеріне де өз ықпалын тигізді. Мысалы, осыдан жүз жыл бұрын Жетісу тұрғыны жаңадан шыққан тауар туралы орта есеппен екі мен жарты жыл аралығында хабар алатын болған. Жарты ғасыр бұрын теледидар, радио және баспасөздің қарқынды дамуы арқасында бұл уақыт мөлшері бірнеше сағатқа дейін қысқарды. Цифрландыру заманында бұл көрсеткіш бірнеше минутқа дейін қысқарды. Ақпараттық-коммуникациялық технологиялар арқасында ақпаратты тарату, соның ішінде жарнаманы тарату әжептәуір жылдамдады. Бұл өз кезегінде жарнама актуалдығы уақытының тез қысқаруына әкелді.

Қазіргі кезде жарнама мәтіндерінің құрылымында да трансформация басталғанын байқауға болады. «Дәстүрлі медиа» деген термин шеңберінде қарастырылатын ақпарат тарату құралдарының мәтінінде көпшілік танып білетін «жарнама» элементтері көрініс беріп жатқанымен, «дәстүрлі емес медиада» мәтін трансформациясы жылдам жүріп жатқанын байқаймыз. Аталған трансформацияны түсіндіру үшін қос жарнамаға тән айырым белгілерін қарастырайық. «Дәстүрлі медиаға» жататын жарнама мәтіндерінде ақпаратты жүйелі беру басты орында болса, «дәстүрлі емес медиаға» жататын мәтіндерде ақпараттың эмоция тудыру аспектісіне ден қояды. «Дәстүрлі медиа» жарнама мәтінінде тауар немесе қызметке тән ерекшеліктер сипатталады. Ал «дәстүрлі емес медиада» тауар немесе қызмет көрсету түріне тән ерекшеліктерін сипаттаумен қатар тұтынушының талғамына, қажеттіліктеріне қарай бағдарланады. «Дәстүрлі медиаға» тән мәтіндерде ақпарат бейтарап түрде жеткізілетін болса, «дәстүрлі емес медиа» мәтіндері тұтынушыны өзінше бір диалогқа шақырады. «Дәстүрлі медиа» жарнама мәтіндерінде эмоцияны білдіретін символдарды пайдаланбай, эмоцияны сөз арқылы жеткізуге тырысса, «дәстүрлі емес медиа» эмоция білдіретін символдарды қолданудан еш қашпайды. Керісінше, «дәстүрлі емес» медиа тұтынушыда эмоциялардың тууына ықпал ететін құралдарды барынша қолдануға мүдделі. «Дәстүрлі медиада» тұтынушымен кері байланыс көпшілікке танымал коммуникация құралдарын (телефон, сайт, мекенжай) пайдаланса, «дәстүрлі емес медиада» кері байланыстың үлкен таңдауын ұсына алады. «Дәстүрлі емес медиадағы» жарнама мәтіндерінде кері байланыс мүмкіндіктері нарықта бар технологиялық шешімдермен шектелген.

Жоғарыда аталған құбылыстар қазіргі замандағы жарнама нарығында жүріп жатқан өзгерістерді сипаттайды. Бұл құбылыстар еліміздегі қазақтілді жарнама нарығында байқалады. Қазақтілді жарнама мәтіндеріне жаһандық құбылыстардың едәуір әсері болып жатыр. Алайда қазақ тіліндегі жарнамаға қатысты мәселелердің де бар екендігін мойындаған жөн. Атап айтқанда, қазіргі кезде қазақ тіліндегі жарнама көп жағдайда аударма жарнама түрінде көрініс тауып келеді. Мысалдарымыз көрсеткендей жарнама мәтіндерінің көбі түпнұсқа күйінде қазақша әзірленбейді. Көп жағдайда дәнекер тіл арқылы аударма түрінде, жиірек түпнұсқа тілден аударма түрінде, сирек жағдайларда қазақтілді тұтынушыға бейімдеген жарнама түрінде жарнама мәтіндерін кездестіреміз. Бұл қазақ тіліндегі жарнаманың сапасына кері әсер етіп келеді.

Сонымен қатар қазақтілді тұтынушының санасында белгілі бір бренд немесе өнімге қатысты теріс түсініктің тууына себепші болып жатқанын мойындау керек. Басқа жағынан аударылған жарнама мәтіндерінің тұтынушыға қате ақпарат беруі әбден мүмкін екенін ескерсек, қазіргі таңда қазақ тіліндегі қате жарнама мәтіндері қоғам үшін өзекті мәселеге айналды. Осы ретте 2021 жылы ҚР Президенті қазақ тіліндегі жарнама мәселесін, атап айтқанда, қазақша аударма жарнамасын ретке келтіру мақсатында «Қазақстан Республикасындағы Тіл туралы» заңға өзгертулер енгізген болатын. Заңның 1-бабына «тентүпнұсқалы аударма» ұғымын енгізді. Аталған бап түпнұсқа тілден немесе дәнекер тілден қазақ тіліне тікелей аудару, сөзбе сөз аудару, сөздің мағынасын сақтамай аудару секілді мәселелердің алдын алуға бағытталған. Алайда бұл норманың жарнама нарығында толық жұмыс жасай бастауы үшін уақыт қажет екенін және сала мамандарының дайындықтан өтуі қажет екенін ұғып отырмыз. Осы жағдайды ескере отырып, қазақ тіліндегі жарнама мәтіндерінің корпусын жасақтау жобасын бастаған едік. Бұл жоба қазақ тіліндегі жарнаманың тұтынушылары үшін көмекші құрал болады деген ойдамыз. Сонымен қатар жарнама саласы мамандары үшін қазақтілді тұтынушылармен тиімді коммуникация құруға жол ашатын, көп мәселенің шешімін табатын қабілеті бар жоба деп есептейміз.

10.4 Жарнама корпусы: сауда жарнамасындағы мәтіндердің жалпы сипаттамалары

Жарнама деп – белгілі бір зат пен құбылысты, іс-әрекет пен қызметті елге танытуын, насихаттауын, оларға деген сұранысты күшейту мақсатымен жарияланған хабарлама жиынтығын айтамыз. Жарнама («жар» – жария етті, мәлімдеді, хабар таратты, жарияланды деген ұғымды берсе; «нама» – араб, парсыда «жазылған хат, шығарылған бәйіт, дастан» – қазақ мәдениетінде кең қолданылған ұғым, дефинициясы айқын термин. Энциклопедиялық, лексикологиялық, мағлұматнамалық, терминологиялық еңбектерде берілген анықтамаларда «жарнама – спектакль, концерт немесе көрермендерге арналған басқа да мәдени шаралар, спорттық жарыстар, ойын-сауық туралы хабарландарудың бір түрі» деп те көрсетіледі.

Қазіргі жарнаманың түрлеріне қарап бөлетін болсақ, біз жарнама түрлерін *сыртқы (көше) жарнама, сауда жарнамасы, теле және радио жарнама, көлік жарнамасы, интернет жарнамасы* деп бөліп қарастыруға болады.

Сыртқы жарнама, бұл аты айтып тұрғандай, даладағы баннерлер, билбордтар, видео, экрандар, дисплейдер және тағы сол сияқты. Көше жарнамасының мақсаты – көшеде жүрген адамдардың назарын тауарға аудару немесе тұтынушының қажеттілігін өтеу, тауардың танымалдығын арттыру. Сауда жарнамасы да дәл осылай сауда орындарының сыртқы және ішкі жағындағы баннерлер, билбордтар, видео, экрандар, дисплейлер және т.б. Жарнаманың таратылу орны мен әдісіне қарай бірнеше топқа бөлініп жұмыс жасалды.

1. БАҚ-тағы жарнама
2. Сыртқы жарнама;
3. Көліктегі жарнама;

4. Сауда орындарындағы жарнама;

5. Интернеттегі жарнама;

Біздің міндетіміз сауда орындарындағы жарнама мәтіндерін жинап, жарнаманың түрлерін анықтау болды:

- сауда залдарын жарнамалық безендірулер;
- дүкен сөрелері (сыртқы және ішкі);
- сауда залындағы белгілер, маңдайша жазулар, планшеттер;
- витрина жапсырмалары; еден жапсырмалары;
- дисплейлер мен теледидар мониторлары;
- жарнама берушінің атаулары мен сауда белгілері жазылған қаптама: қораптар, футлярлар, орауышқағаз, скотч және т. б.;
- манекендер мен бутафория;
- бликфанг;
- слайдтар мен сөрелер;
- фирмалық баға белгілері;
- воблерлер; шелфтокерлер;
- кассалық аймақтағы жарнама;
- чекпоинт;
- бөлшек сауда ақпараттық жүйелері арқылы ауызша жарнама.

Жұмыс барысында, жарнама жинау мәселесін қайдан және қалай жинау мәселесі тұрды. Ең алдымен Алматы қаласындағы сауда орталықтарын іріктеп алдық. Бірінші кезеңде ең ірі сауда орталықтары, екінші кезеңде Алматы қаласындағы барлық базарлар қамтылды. Үшінші кезеңде шағын дүкендер алынды. Осы аталған орталықтардағы түрлі жарнамалар суретке түсіріліп, жинақталды:

Кесме 21

Сауда орталықтарындағы жарнамалардың метабелгіленімі

Жарнама материалы	Жарнама мәтіні	Жарнама түрі	Жарнама тасырыбы	Жарнама формасы	Жарнама тілі	Жарнама жылы	жарнама беруші	Берілу аймағы	
	Сиритэй Самал	Көшедегі жарнама	Мақро топ	минералды су	Визуал	Билборд	2023	Ірі бизнес	Алматы қаласы, Бузово ауданы
	LAVIVA ULKER Талдыбұлат ауданының аумағында	Көшедегі жарнама	Сусын	шоколад	Визуал	Билборд	2023	Ірі бизнес	Алматы қаласы, Бузово ауданы
	Мәңгілік елдің адам	Көшедегі жарнама	Әлеуметтік	дәуіс беру	Визуал	Билборд	2023	Ірі бизнес	Алматы қаласы, Бузово ауданы
	Ойыңа ұяда қуат жина	Көшедегі жарнама	Сусын	тәтті газдалған сусын	Визуал	Билборд	2023	Ірі бизнес	Алматы қаласы, Бузово ауданы
	Аккумулятор сату орнату кендеу желісі	Көшедегі жарнама	Техника	орталық	Визуал	Билборд	2023	Ірі бизнес	Алматы қаласы, Бостандық ауданы
	MacCoffee сервисіміз Di Torino Италиялық сепілі бар кофе	Көшедегі жарнама	Сауда	дүкен	Визуал	Билборд	2023	Орта бизнес	Алматы қаласы, Бостандық ауданы
	Дәріс ұйықтау үшін диван-көреуіттерге 65% АЛКОМА.КЗ	Көшедегі жарнама	Тұрмыстық заттар	Үй жиһазы	Визуал	Билборд	2023	Орта бизнес	Алматы қаласы, Бостандық ауданы

Алматы қаласы бойынша қарастырылған сауда орталықтарының нәтижесінде байқағанымыз: қазақша жарнамаға қарағанда, орысша жарнама көбірек болды.

Қазақша және орыс тіліндегі жарнамалардың пайыздық көрсеткіші

<i>Сауда орындарының түрлері</i>	<i>Қазақша жарнама</i>	<i>Орысша жарнама</i>
Құрлыс материалдарға арналған сауда орындар: Барлық, Кар Сити, Семиречье	20 %	80 %
Үлкен сауда орталықтар: Москва, АДК, Променад, Гранд Парк, Мега Центр	40 %	60 %
Шағын сауда орталықтар (дүкендер): Восход, Тиграхауд	40 %	60 %
Орталықта орналасқан сауда орындар: Цум, Есентай Молл, Достық Плаза	40 %	80 %
Шеткі аудандарда орналасқан сауда орындар Апорт, Алматы Молл	30 %	40 %
Базарлар (Адем, Байсаг, Рахат, Алағау, Арлан, Ақ-бұлақ, Кенжехан	20 %	30 %

Жарнамалық хабарламаны біз ауызша және визуалды семантикалық компоненттердің бірлігі ретінде қарастырамыз:

- хабарламаның вербалды компоненті деп әдетте мыналар түсініледі: аты (фирма, бренд, жарнама объектісі), сөз (мәтіннің құрылыс материалы), ұран;
- сөз вербалды емес немесе визуалды компоненттер деп кескін (ақпарат, шифрланған және бейнелеу қатарында берілген) және композиция (сурет пен мәтіннің үйлесімі) түсініледі.

Жарнамалық хабарламаны бейімдеудің негізгі тетіктерінің бірі ауызша және вербалды емес семантикалық компоненттер арқылы жасалады және компанияның, сондай-ақ тұтынушының жарнамалық бейнесі түрінде ұсынылады. Кескіннің өзі өте шектеулі уақыт пен кеңістікте өзекті болып қала береді.

Д.В.Беклешов және басқа да мамандар бейімделу тек жарнамалық хабарлама визуалды және ауызша құралдармен, психологиялық, мәдени, демографиялық, заңнамалық және ұлттық бейімделу стратегиясының басымдықтарымен қатар көрсетілген талаптардың жиынтығы болған кезде ғана тиімді болады деп санайды [295, 85 б.].

А.В.Овруцкийдің айтуынша, жарнамалық сурет – бұл тек жарнамаланған тауармен, брендпен байланысты мозаикалық құрылым және олардың өмірі әрдайым басқа бейнелермен, сондай-ақ жарнама берушінің бейімделу және маркетингтік науқанының мақсаттары мен стратегияларымен шектеледі.

Жарнамалық бейненің әсерінен көрінетін бейімделу механизмі кез келген жарнамалық хабарламаларды соншалықты таныс және қарапайым етуге мүм-

кіндік береді, олар ерте ме, кеш пе, олар ұжымдық сана мен жарнаманы тұтынушылардың бейсаналық ойына тікелей кіреді. Кескін жарнамалық хабарлама бейімделген елдің тұтынушысының мәдени құндылықтарына сәйкес келуі керек [296, 95 б.].

Қазақстандағы жарнама мәтіндері ауқымды және әсерлі деп есептейміз. Дегенмен, қазіргі уақытта ол зерттеліп, дұрыс бағалануы тиіс. Қазіргі кезде батыстық мамандар әзірлеген әдістемелерді, тіпті дайын жарнама мәтіндері кеңінен қолданылууда.

Жарнама мәтіндері ең маңызды коммуникативті функцияны орындайды. Аудармашы түпнұсқа мен аударманың мазмұндық арақатынасын жеткізіп қана қоймай, оның әлеуметтік лингвистикалық ерекшеліктерін де жеткізуі қажет.

Жарнама мәтіні тек лингвистикалық, сөздік компонентті ғана емес, сонымен қатар бірқатар экстралингвистикалық компоненттерді де қамтиды. Және олар лингвистикалық және экстралингвистикалық белгілердің үйлесімді болған жағдайда ғана қабылданады.

Экстралингвистикалық факторлар ұлттық тілдік суреттің ерекшеліктерін де қамтиды. Әрбір тіл шындықты өзінше бейнелейді, сондықтан тілдер «әлемнің тілдік суреттерімен» ерекшеленеді. Белгілі әдебиеттанушы және когнитивист В.А.Маслова жазғандай, «әлемнің лингвистикалық суреті ішінара әмбебап, ішінара ұлттық болады. Сондықтан оның ұлттық ерекшелігін әртүрлі халықтар әлемінің суреттерін салыстыру арқылы ғана жүзеге асыруға болады».

Жарнамалық мәтінді лингвомәдени талдаумен қатар жарнамалық хабарламаларда бірнеше бағыттар бар:

1. Маркетингтік ақпаратты аудару.

Егер мақсатты аудитория өкілдерінің мінез-құлық портреті түпнұсқаға ұқсас болса, онда бірінші деңгейдегі бейімделу жаңа нарықта сәтті алға жылжу үшін жеткілікті болады. «*Джиллетт – ер адам үшін жақсы емес*» жарнамалық хабарламасының мысалын қарастырайық. Ағылшын тілінде берілуі: «*Gillettthebest a mencanget!*». Бұл қазақ тіліндегі аудармада бастапқы дереккөздің стилі мен мазмұны, сондай-ақ мәтіндік хабарламаның мөлшері мен екпіні сақталған. Бұл мысал хабарламаның семантикалық және маркетингтік ақпаратының сапалы аудармасын көрсетеді;

2. Жарнамалық хабарламаның ауызша немесе визуалды компонентіне ұлттық сипат беру.

Бұл қарапайым аударма емес, жарнаманы оқшаулау деп аталатын шығармашылық процесс болып табылады. Лексикалық нюанстардан басқа, мақсатты аудиторияның әлеуметтік-мәдени, құқықтық және моральдық контексті ескеріледі. Бұл ретте жолдаудың жалпы ақпараты сақталады.

Мысалы, теледидарлық жарнама *SNICKERS* нұсқасында «*Сен аш болсаң, өзіңе ұқсамайсың*» әлсізді мықтыға айналдырады. Ал бұл Қазақстанға келгенде басқа роликке ауысуы мүмкін. Бұл әр елдің менталитетіне байланысты.

3. Әртүрлі нарықтарға бағытталған әртүрлі жарнамалық науқандар мен хабарламалар жасау. Жаңа жылдық жарнамалар, 8 наурыз жарнамалары және Наурыз мерекесіндегі жарнама мәтіндері жатады.

Әр ел үшін өз роликтері және жарнамалары жасалады. Жаңа мақсатты нарықтағы маркетингтік бөлімшенің міндеті – жарнамалық хабарламалар үшін ең қолайлы сценарийді таңдау болып табылады. Осы қағидатқа байланысты жарнамалық идеяны жүзеге асыруға шығармашылық көзқарас көбінесе нарықтың жергілікті ерекшеліктеріне, тұтынушының тіліне, оның сұраныстарына және т. б. бейімделіп өзгеруі керек деген қорытынды жасауға болады.

Жарнама мәтінінің түрлерін толығырақ қарастырайық:

1. Жарнаманың вербалды коммуникативті түрінде негізгі қатынас құралы сөз болып табылады, ал бұл жағдайда қарым-қатынас формасы жазылады.

2. Вербалды-бейнелік коммуникативті типпен сөз арқылы білдірілген семантика көрнекі компоненттермен толықтырылады. Бұл жерде рационалды немесе эмоционалды бейнелеудің табиғатын атап өту керек, ол бірінші жағдайда тек денотат белгілерімен анықталады; мағынаны елестету кезінде бейнелеудің бейнелілігіне оның көрнекілігінің арқасында қол жеткізіледі.

3. Жарнаманың аудио-вербалды коммуникативті түрінде мәтін құрылымының күрделенуі жаңа коммуникация арнасының, атап айтқанда сөйлеу мен тыңдаудың қосылуына байланысты.

Жарнаманың сөздік бөлігінің құрылымдық элементтеріне 4 негізгі бөлік кіреді:

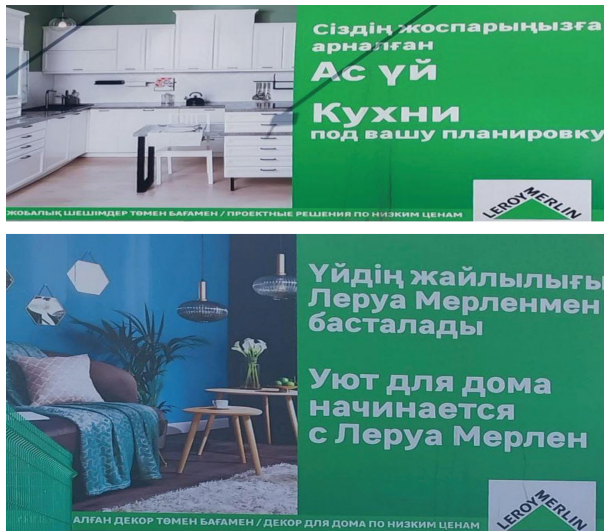
- 1) ұран;
- 2) атауы;
- 3) негізгі жарнама мәтіні;
- 4) қайталау сөйлем.

Ең маңызды тақырып функциялары:

- а) назар аудару
- ә) қызығушылықты ояту;
- в) сатып алушыны/мақсатты топты анықтау;
- г) өнімді/қызметті анықтау;
- д) өнімді/қызметті сату [297, 51 б.].

Жоғарыда аталған сауда орындарынан материал жинау барысында түрлі жарнамаға кез болдық. Сауда орталықтарының ішінде ерекше жарнамасы толық қазақшаланған «Леруа Мерлен» сауда орталығын атап өтуге болады. Орталыққа беттеген тақтайшадан бастап сауатты, нақты, дәл жарнамаға тап боласыз. «Леруа Мерлен» орталығының негізін қалаушылар шетел азаматтары яғни бұл – франшиза. Отандық сауда айналымдағы әлемдік танымал брендтердің барлығы (оның ішінде қазақстандық маркалар да бар) сауда белгілері: тұтыну тауарлары, киім-кешек, автокөліктер, азық-түлік, тіпті компьютер ойындары мен кинофильмдер франшизаға жатады. Бір сөзбен айтқанда, келісімшарт арқылы жергілікті кәсіпкерлер шетелдік және отандық өнімді отандық нарықта сата алады.

Бұл кісілер бизнесті аса меңгерген бе, әлде шетел азаматтарына талап жоғары болды ма? Қысқасы, Алматы қаласы бойынша өте сапалы жарнаманы тек осы орталықтан кездестірдік:



Мысалы, «Үйдің жайлылығы Леруа Мерленмен басталады». «Сіздің жоспарыңызға арналған АС ҮЙ». «Леруа Мерленнің қолдауымен өз бақшаңызда өсіріңіз». Көшеттер, топырақтар және тыңайтқыштар төмен бағамен. Гүл құмырасы. Өсімдіктерді ылғалдандырғыш. Тыңайтқыштар. Москит торы. Қоқыс қаптары. Көгал тұқымы. Қылқан жапырақты тыңайтқыштар т.б.

Жарнамалық хабарламадағы жарнамалық визуалды сурет тұтынушыға жарнамаланатын өнім немесе қызмет туралы жаңа нәрсені анықтауға мүмкіндік береді. Жарнамалық хабарламалардың көркемдік дизайны жарнамалық идеяны бейімдеудің негізгі компоненті ретінде түрлі түсті кескіндер мен иллюстрацияларды қолдануды қамтиды. Жарнама объектісі мен көркемдік дизайн арасында көрерменге түсінікті логикалық байланыс болуы керек, осыған байланысты бейімделу процесі жүреді. Көрнекі материал көпшілікке қызықты және түсінікті болуы керек, ал фактілер дәл және сенімді болуы керек.

Жарнамалық хабарламаны жасау үшін қоғамның басым тобының идеясын жүзеге асыратын стереотиптік механизм қолданылады.

Ғылыми әдебиеттерде айтылғандай, «стереотип – шартты рефлекторлық әрекеттердің бекітілген стереотиптік тәртібі түрінде сана мен мінез-құлықта көрінетін мидың тұтас, жүйелі әрекетінің нысаны» болып табылады.

Көптеген ірі халықаралық компаниялардың Қазақстанда өз өкілдіктері бар және қазақстандық тұтынушыға өз өнімдерін ұсынады. Бұл дегеніміз, осы өнімдерді жылжыту науқаны аясында трансұлттық корпорациялардың халықаралық жарнамасы отандық нарықта пайда болады. Жарнамалық хабарламалар әлеуметтік-мәдени алғышарттарға негізделеді, адами байланыстардың әртүрлі салаларына – идеологияға, дінге, өмірге, эстетикалық қызметке, билік ұмтылыстарына және т.б. әсер етеді. Қазақстан нарығына бағытталған жарнамалық хабарлама кез келген ішкі нарыққа тән нақты әлеуметтік-мәдени ерекшеліктерге байланысты ішінара немесе толық мәдени бейімделуді қажет етеді.

Жоғарыда айтылғандарға сүйене отырып, 1) Табысы бірдей болғандықтан, халықтың әртүрлі топтары басқа сипаттамалары бойынша бір-бірінен айтарлықтай ерекшеленеді. Мысалы, қаражаты аз жастар мен қарт адамдар да кедей. Дегенмен, тауарларды, қызметтерді, тіпті одан да көп мәдени құндылықтарды тұтытуда бұл топтар арасындағы айырмашылықтар түбегейлі;

2) халықтың белгілі бір топтарының психографиялық сипаттамалары;

3) тұтынуға елеулі әсер ететін келесі фактор елде құрылған инфрақұрылым болып табылады, оған дистрибуция жүйелері, көлік коммуникациялары, интернетті, телефонды пайдалану мүмкіндіктері, автокөлік құралдарының болуы кіреді;

4) мәдени дәстүрлер. Кейбір қазақстандық популяцияларға шынайы американдық құндылықтарға негізделген тауарлар мен қызметтерді ұсыну күлкілі болар еді.

Тұтынуға әсер ететін кішігірім факторларға географиялық ерекшеліктер, климат, халықтың ұтқырлығы, белгілі бір елдегі немесе аймақтағы нақты жағдайлар және тағы басқалар жатады.

Осылайша, бір жағынан, жарнама ұлттың рухани мәдениетінің қоймаларының бірі, ал екінші жағынан, мәдениетаралық қарым-қатынас процесінде ұлттық мәдениетті танымал етудің күшті факторы болып табылады.

Бұл ретте, өзге мемлекеттегі әлеуетті тұтынушыға жарнамалық жолдаудың мәнін дұрыс жеткізу үшін бірінші кезекте жарнамалық жолдаудың орындылығын ескеру қажет. Жарнама беруде жарнамалық ұранның негізгі идеясын сақтауымыз керек.

10.5 Жарнамада берілген имплицитті мәтіндерді түсіну ерекшелігі

Жарнама мәтіннің прагматикалық әлеуеті оны адамның қаншалықты жақсы қабылдап, қаншалықты жақсы түсінгеніне және мәтіннің қоғам санасында интерпретациялануына тікелей байланысты.

Мәтінді түсіну мәселесі, яғни «хабарды жеткізушінің» интенциясын анықтау бағыттары ұзақ уақыт бойы риторика мен герменевтика пәндерінің зерттеу еншісінде болды [298, 87]. Герменевтика (гр. *Hermeia* – ұғындыру, баяндау, түсіндіру) – кең мағынада философия мен әлеуметтік болмыстың мәнін түсінудің шарты ретінде қарастырылатын бағыт. Тар мағынада – білімнің біраз салаларында (философия, әдебиет, тарих, филология, дінтану т.б) мәтінді түсіндірудің, ұғындырудың ережелері мен техникаларының жиынтығы [299, 68]. Бұл жерде біз герменевтиканы соңғы мағынасында қолданып тұрмыз.

Жарнама мәтіндерінің түсінілуін зерттеудің жаңа мүмкіндіктерін соңғы онжылдықта лингвистика ғылымында пайда болған жаңа бағыт когнитология жүзеге асырып келеді. Когнитология (лат *cognition* – білім, түсінік, көзқарас) – тіл біліміндегі бағыт, тілді игеру, табиғи тіл механизмін түсіндіру үшін білім жинау, қолдану және оны түсіну моделін құру тұрғысынан зерттеу, яғни тілдің танымдық теориясы мен танымдық қызметін зерттеп, адамзат білімінің түзілуі мен оның қызмет болмысын қарастырады [94, 80]. Қазіргі уақытта, танымдық

процестер мен механизмдердің көмегінсіз жарнама мәтіндері мен креолданған мәтіндерді түсіну, яғни олардың санада көрініс табуының лингвистикалық мәселелерінің шешімін табу мүмкін емес. Жалпы мағынасында түсіну үдерісі когнитология ғылымының зерттеу нысаны болып табылады, оның негізгі мақсаты зерттеуге алынған объектінің санадағы мәнін белгілеу [300, 9]. Мәтінге қатысты түсіну ұғымы перцептивті-танымдық-аффективтік және қабылданатынды өңдеуді қамтитын кешенді көпсатылы үдеріс. Бұл білімнің тілдік және энциклопедиялық түрлерінің өзара қатынас жасауын қажет етеді.

Жарнама мәтіндерін қабылдау мен түсіну процесі мерзімі жағынан қысқа, яғни өту уақыты шектеулі болатындығымен ерекшеленеді. Бұл жағдай жарнама коммуникациясының өту жағдаятына және оның сипаттамалары мен шарттарына байланысты түсіндіріледі. Сондықтан тиімді жарнама мәтінін құру талабы бойынша копирайтерлер қысқа хабарламада ақпараттың максималды көлемін беруге ұмтылады, сонымен қатар хабарлама формасының ерекшелігін сақтауға да тырысады. Бұл процесс «минимум қаржы, максимум ақпарат» ұстанымы бойынша жүзеге асады. Осының салдарынан ақпараттың жарты бөлігі таңбалармен белгіленбей қалады, алайда адресант арқылы тұспалданып тұрады, яғни имплицитті түрде беріледі. Имплицитті (лат. *implicite* жасырын, анық емес) – тілдік материал арқылы ашып көрсетілмеген, бірақ контекст арқылы білініп тұратын жасырын мағына, бір сөзбен айтқанда, имплициттеу дегеніміз – тұспалдау, бір нәрсені тілдік құралдар арқылы тікелей емес, жанамадай көрсету, меңзеу [301, 337-369]. Мысалы, «Тиімді бағадағы премиум кластық үй» деген жарнаманы алатын болсақ, бұл жарнама «жоғары сапалы, жайлы өмір сүруге қажетті нәрселердің барлығымен жабдықталған, отандық танымал құрылыс компаниясы салған үй» дегенді меңзейді. Жарнамадағы бұл ақпаратты «премиум кластық» деген анықтауыштық тіркес арқалап тұр. Мұндай тұспалды жарнама тұтынушылары дұрыс түсінуі де мүмкін немесе оған мән бермеуі де ықтимал. Осыдан барып имплициттілікті «айтылмайтын», алайда «тұспалданатын» деп анықтайтын тұжырымның көңілге қонымдылығы байқалады. Ю.К. Пирогова жарнама коммуникациясына ықпал ететін имплицитті ақпараттың негізгі үш ерекшелігін көрсетеді:

1) жарнамада берілген нақты ақпараттан имплицитті ақпараттың басты айырмашылығы – имплицитті ақпарат адресаттың түйсінуінен тыс, яғни оның ойлау шеңберін айналып өтіп, ақпаратты өңдеу үдерісінен тысқары жүзеге асатын құбылыс. Сондықтан адресат мұндай ақпаратты бағалауға, оған күмән келтіруге бейім емес.

2) адресат мұндай ақпаратты дайын күйінде емес, хабарламаны интерпретациялау кезінде өзі шығарады, сондықтан оны қаз-қалпында, еш өзгеріссіз қабылдайды.

3) имплицитті ақпаратты пайдаланудың үшінші себебі жарнама қызметінің құқықтық аспектісімен байланысты. Жарнама берушіні жарнамада анық берілмеген ақпарат үшін жауапқа тарту (жарнама қызметіне байланысты заңнамаға сәйкес) мүмкін емес [302, 86-88].

Д.Р.Теркулова жарнама дискурсындағы имплицитті ақпараттың берілу ерекшеліктеріне кешенді зерттеу жүргізді. Бұл зерттеуде жарнама мәтнінде

тілдік ерекшеліктерді жүзеге асырудың жанама түрлері импликауралар қарастырылды:

1) теориялық импликауралар дүниенің бейнесі туралы жалпы білімге негізделеді;

2) дәстүрлі (конвенциональные) импликауралар дүниенің тілдік бейнесі туралы білімге негізделеді;

3) коммуникативті импликауралар ауызша қарым-қатынас ережелеріне сүйенеді.

Мәселен, жоғарыда біз беріп өткен «Тиімді бағадағы премиум кластық үй» деген жарнама теориялық импликауралардың көмегімен батыс мәдениетінің құндылықтарын жетік білетін және оны ұстанатын белгілі бір әлеуеттік топтың дүниенің бейнесі туралы жалпы біліміне негізделіп берілген жарнама болып табылады. Себебі, мұндай топтың санасында «премиум кластық үй» деген анықтауыштық тіркес «жақсы өмір, өзің үшін өмір сүр, жақсы сервис т.б.» секілді концептілік білімдер жүйесін құрайды. Ал егер де біз осы жарнаманы дәстүрлі импликауралардың көмегімен, дүниенің тілдік бейнесіне сүйеніп беретін болсақ, онда ол: «Тиімді бағадағы сенің отбасыңа арналған үй» деп басқаша көрініс табар еді. Себебі, қазақтың дәстүрлі санасында дүниенің тілдік бейнесі туралы жалпы білімінде «үй» концептісі «отбасы, бірлік, бала-шаға т.б.» ұғымдарынан тұрады.

А.Н. Баранов импликаураларды «айқын көрсетілмеген нәрсе» деп анықтайды. Кейбір лингвистердің еңбектерінде импликауралар тек бір жақты түсіндіріледі. А.Н. Баранов сөздігі бойынша, импликаура – сөйленім мағынасының бір бөлігі. Ол айтылған нәрсенің сөзбе-сөз мағынасынан бөлініп шығарылады [303, 647]. Г.П.Грайстың айтуынша, импликауралар деп имплициттенген нәрсені түсінеміз [304, 52-58]. Г.Г.Молчанова имплицитты түсіндіруде логикалық терминдерге сүйенеді. Қазіргі таңда когнитивті лингвистиканың аясында, когнитивті семиотика мен когнитивті модельдеу терминологиясы бойынша біз, имплицитты жарнама мәтнінде имплициттенген ақпараттың қамтылғанын көрсететін белгілер екенін түсінеміз. Имплициттар ретінде адресаттың санасында берілген білімдер жүйесін анықтайтын креолданған жарнама мәтнінің вербалды және бейвербалды, индекстелген элементтері қолданылады. Г.Г.Молчанова импликаураларды логикалық деңгейде вербалды түрде көрініс беретін имплициттенген мәтіннің жасырын деңгейін көрсететін шартты бірлігі ретінде қарастырды. Оның ойынша, импликауралар жадыда іске қосылатын фреймдерді танудың импульсі болып табылады. Фрейм – санада концепт түзеуші оқиға бірліктерінің бірінен кейін бірінің жүйелі, тәртіпке сай өтілуін қадағалайтын ақпараттар жинағы, концептілік жүйенің құрылымдық элементтерін танытатын қарапайым нобай [305, 6]. Фрейм жарнама мәтнінің интерпретациялану және инференциялану процестерін моделдеуге мүмкіндік туғызады. ЖАРНАМА ОБЪЕКТІСІ ретінде санада танылатын фреймдер жарнама мәтні туралы ақпарат беретін «тұтынушы, өнімнің мінездемесі, өндіруші туралы ақпарат, мақсатты аудитория» т.б. сияқты бірнеше элементтерден тұрады. Сондықтан жарнама мәтнінде фрейм эксплицитті және имплицитті формаларда беріледі. Мысалы, «Арманыңыздан

асқақ әлем Анталияға ұшамыз» деген жарнама мәтінде негізгі имплицат – Анталия, бұл элемент мәтінде инференцияға ұшырайтын имплицитті формадағы фреймді білдіреді. Ал, «Арманыңыздан асқақ әлемге ұшамыз» деген вербалды компоненттер мен әуе тасымалдаушы компанияның логотипі қосалқы имплицаттар болып табылады. Сонымен берілген жарнама мәтіні «сенімді әуе тасымалдаушы компаниясының көмегіне жүгініңіз, EkonomPlus компаниясының билеттерін сатып алыңыз» деген имплицитті мағынаны береді.

Сонымен, жарнама мәтінін түсіну оның тек анық берілген мағыналарын оқумен ғана шектелмейді, сондай-ақ оның құрылымының айтарлықтай бөлігін инференция арқылы ажыратып, толық мағынаны сол арқылы қалпына келтірумен жүзеге асады. Е.С.Кубрякованың айтуы бойынша кез келген тілдік форма, оның ішінде мәтін өзінде анық барды ғана емес, сонымен қатар астыртын семантикалық мағынаны, яғни инференция құбылысын да береді [306, 256-260]. Инференция ой процесі ретінде түсіндіріледі. Инференция кезінде адам өзі қарастырып отырған тілдік формалардың, тілдік бірліктердің дәлме-дәл, сөзбе-сөз мағынасынан шығып, оның негізгі мағынасынан бөлек, тереңірек мазмұнын анықтай алады [307, 53-61]. Инференция құбылысы жарнама коммуникациясы барысында адамның адресант пен адресаттың аражігін ажырата алатындай базалық білімі болған жағдайда ғана жүзеге асады. Т.В.Дроздова өзінің зерттеу жұмысында белгілі бір білім түрлерінің инференцияның нақты бір формасымен сәйкестігінің матрицасын құруға талпынған. Алайда бұлай жасау мүмкін болмай шықты, себебі инференция құбылысында бірін-бірі өзара толықтыратын білімнің бірнеше түрі жүзеге асады [308, 217-237].

Инференция процесі кезінде жарнама мәтінде инферентті ақпараттардың келесі түрлері беріледі:

- жарнама объектісінің қызметі туралы;
- жарнама өндіруші туралы;
- жарнама объектісінің ерекшелігі оның мінездемесі туралы;
- жарнама тұтынушысы туралы.

Осы фреймдік ақпараттардың көмегімен имплицаттар өзара байланысқа түсіп инференция құбылысы жүзеге асады. Әртүрлі имплицаттардың өзара байланысуы арқылы жүзеге асатын инференция процессінің ерекшеліктерін қарастырса, жарнама объектісі белгілі бір қызметтер жүйесінен тұратыны, оның қызметі жөніндегі ақпарат имплицитті түрде берілетіні айқын байқалады. Мысалы,

1. Мезим. Асқазанға таптырмас ем! (Мезим для желудка незаменим!)
2. Аллергия ма? Шешімі бар – Кларотодин! (Аллергия? Есть выход – Кларотодин!)
3. Тежелме, сникерстен! (Не тормози сникерсни!)

Бірінші, екінші мысалдардың негізгі имплицаттары Мезим мен Кларотодин, бұл екі имплицат жарнама мәтінінің интерпретациялануының негізі болып табылатын жарнама объектісін құрайды. Жарнаманың қосалқы имплицаттары асқазан және аллергия жарнама мәтінде вербалды түрде берілген, олар жарнамаланып отыраған дәрілік құралдардың қолданыс аясын көрсетеді. Имплицитті ақпараттардың түпкі мағынасы имплицаттар арасында олардың

қолданыс аясын меңзеп тұратын концептуалды байланыс орнаған кезде анықталады: Мезим – асқазан, Кларотадин – аллергия. Мұндай концептуалды байланыс негізгі импликактар мен қосалқы импликактар арасында олардың мәтіндегі қызметтеріне сәйкес орнайды. Сонымен, бірінші жарнама мәтіні «Мезим асқазан ауруы кезінде ауруды басатын емдік құрал», ал екіншісі – «Кларотадин аллергиядан құтылуға көмектесетін дәрі» деген инференцияланған ақпаратты беріп тұр. Ал үшінші жарнамада негізгі импликакт рөлін «тежелме (не тормози)» сөзі атқарып тұр. Себебі бұл жарнаманың мақсатты аудиториясы – жасөспірім балалар. Жарнама ең алдымен солардың тілінде сөйлеп, «сникерс жесең кешке дейін тоқ жүресің» деген астыртын мағынаны беріп тұр.

Сонымен, қорытындылайтын болсақ, инференция құбылысын анықтауға арналған бүгінгі қолданыстағы тәсілдер бізге жарнама мәтінінің инференциялануының анықтамасын қисынға келтіруге мүмкіндік берді. Инференция – бұл адресаттың жарнама мәтінінің вербалды түрде және белгілі құралдар арқылы берілмеген мазмұнын ойша қалпына келтіру операциясы. Ол санада жарнама мәтіні элементтерін жаңартуға негіз болатын белгілі бір құрылым түрінде сақталған білімге негізделеді. Қазіргі кезде зерттеулердің көптеп жүргізіліп жатқандығына қарамастан, тілтаным тарихында бұл құбылыс өте аз зерттелген құбылыстар қатарына жатқызылады.

11 ЛИНГВИСТИКАЛЫҚ БЕЛГІЛЕНІМ ӘЗІРЛЕМЕЛЕРІ

11.1 Дерексіз зат есімдер: Қазақ тілінің ұлттық корпусында лексикалық белгіленімді анықтаудың кейбір мәселелері

Қазақ тілтанымында, әлемдік лингвистикада да грамматикалық категориялардың қатарында мейлінше назар аударылған қабат – зат есім деуге болады. Жалпы тіл дегеніміз қарым-қатынастың басты құралы деген аксиомаға тоқталмағанның өзінде, тұтастай алғанда, тіл қоршаған ортаны бейнелеудің, ақиқат дүние туралы білімнің сақталуының, қоршаған орта мен ақиқат дүние жөнінде жалпы білім алудың негізгі және бірден-бір формасы екендігі белгілі. Ал бұл ретте қоршаған орта мен ақиқат дүниедегі зат, құбылыс атауларын нақты атайтын сөздердің рөлі зор. Осы себепті лексикология, сөзжасам, синтаксис, стилистика салаларындағы зерттеулерде эмпирикалық дереккөз ретінде зат есім қабатының мағыналық, құрылымдық түрлері – деректі және дерексіз зат есім, дара және жинақтық зат есім, жалпы және жалқы зат есім деп аталатын семантикалық категориялар талданады. Зерттеулерде зат есімнің дербес сөз табы екендігі, қандай сұрақтар арқылы анықталатыны, сөйлемдегі атқаратын қызметі, түрлі топтары қарастырылады. Тұжырымдар жасауда негізінен деректі зат есімдердің сипаттамалары таратылып айтылады. Бұған да негіз бар деуге болады; өйткені нақты зат есімдер сын есіммен, сан есіммен тіркеседі, өзара тіркеседі, сөзжасамның түрлі амалдары арқылы жаңа мағына, жаңа сөз жасауға қатысады, қосымшалар арқылы түрленеді. Бұл – талдаудың құрылымдық жағы. Талдауға алынатын еңбектерде екі ерекшелік негізге алынады: қазақ тілтанымындағы зерттеулерде шетел тілдеріндегі *plan of expression and plan of content* (ағыл.), «план выражения және план содержания» (орыс тілінде) дегендерге сәйкес келетін *сөздің сыртқы тұрпаты, мазмұн тұрпаты* немесе *ішкі мазмұны, тұрпат межесі, мазмұн межесі* сияқты аталымдар қолданылып келеді. Әрине, бұл арада сыртқы тұрпаты деген сөздің әріптер (айтылу кезінде дыбыстар) кешені, жиынтығы, яғни сөздің сыртқы қабығы дегенді білдіретіні анық. Ал мазмұн тұрпаты әріптер кешеніне қарағанда анағұрлым күрделі. Сондықтан тіл ғылымындағы жаңа бағыттардың әдіснамалары тіл қорындағы зат есімдер тек қана грамматикалық категория емес, тілдің нақ осы қабатының ерекшеліктері тілдің төлиесінің когнитив өрісіне, адамның әлемнің бейнесін қалай қабылдап, қалайша жеткізетіндігіне байланысты. Бұл әсіресе дерексіз зат есімдерді зерделеудің өзектілігін байқатып отыр.

Қазақ тілінің Ұлттық корпусы (әрі қарай – Ұлттық корпус) үшін зат есімдердің семантикалық метабелгіленімін анықтау барысында алдымен аңғарылғаны, мәселен, тек АБА – БАРМАҚ сөздері арасында (4200 сөз) дерексіз зат есімдер шамамен 30-40%-ға жуық. Бұл – танымның, биік санадағы ойлаудың айқын көрінісі, өйткені дерексіздік заттар мен құбылыстардың қасиет-сапаларының өзара қарым-қатысы арқылы сол заттар мен құбылыстардың аса мәнді қасиет-сапаларының сипатталуының, нақтылануының нәтижесі болып табылады. Екіншіден, адамның санасында және жадында заттың, ұғымның визуалды образын танытпайтын, белгілі бір бейнені бере алмайтын, алайда тілді

қолданушының қоршаған орта мен ақиқат дүниеге қарым-қатысын білдіретін «дерексіз» (өзара мағыналас – интеграл мағына) болып танылғандардың өзін құрылымдық сипаттамадан шығып, танымдық негізде анықтаудың өзектілігі анық байқалды. Ал танымдық негізде анықтап алу үшін, сөз жоқ, сөздің грамматикалық тұрпатымен қатар лексика-семантикалық мағыналарын (дифференциал мағына) танытудың маңызы зор.

Нысананы талдау және ұсыныстар жасау үшін қазақ тілтанымының көрнекті өкілдері А. Байтұрсынұлының, Е. Омарұлының, Қ. Жұбановтың, Р. Сыздықтың, А. Ысқақовтың, Н. Оралбайдың А. Жұбановтың, Н. Уәлидің, Э. Сүлейменованың, А. Фазылжанның, А. Жанабекованың, сондай-ақ орыс, татар тілдері бойынша бірқатар шетел ғалымдарының зерттеулері қарастырылды. Лексикографиялық еңбектер (Орфографиялық сөздік, 2013; Қазақ әдеби тілінің түсіндірме сөздігі, 2012, Қазақша-орысша 30 томдық терминологиялық сөздік), энциклопедиялық анықтамалықтар мен қазіргі жазба коммуникациядағы тілдік фактілер іріктеліп, топтастырылды.

Ұлт ұстазы – Ахмет Байтұрсынұлы зат есімнің түрлеріне нақты анықтама берген: «Зат есім – нәрселердің атын көрсететін сөздер. Нәрсе де не түрлі де болса заттарды айтамыз. Зат, тегінде, екі түрлі. Біреулері – көзге көрінетін, құлаққа естілетін, мұрынға иіс келетін, тілге татылатын, денеге сезілетін деректі заттар. Екіншілері – көзге көрінбейтін, құлаққа естілмейтін, мұрынға иіс келмейтін, тілге татылмайтын, денеге сезілмейтін, адам ақылмен ғана білетін дерексіз заттар» [309, 199-200]. Осылайша ғалым *жер, су, ай, күн, дауыс, дыбыс, ән, күй, иіс-қоңыс, дәм, ащылық, тұщылық, жылулық, суықтық* тәрізділерді деректі деп, ал *бақ, бәле, жан, өлім, ой, мінез, ұят, реніш, қуаныш* дегендерді дерексіз зат есімдер деп ажыратып көрсеткен.

Зат есімдердің лексика-грамматикалық сипатын профессор, белгілі ғалым А. Ысқақов жан-жақты баяндаған. Ғалымның тұжырымы бойынша: а) абстракті (дерексіз) ұғымдар *көзбен көріп, қолмен ұстауға болмайтын, ойлау, топшылау қабілеті арқылы танылатын* сөздер, олар: *ақыл, сана, ой, ықпал, әсер, жақсылық, адамгершілік, кіріс, зұлымдық, пікір*; ә) дерексіз зат есімдер нақты зат есімдердің негізінде жасалған; б) адамның ойлау қабілеті дами келе, абстракті ұғымдардан конкретті ұғымдар да жасалады; б) осы себепті қазіргі тілдегі абстракті ұғымдарды білдіретін көп сөздердің мағыналары конкретті ұғымдармен түйісіп отырады; в) сөз мағынасының деректену және дерексіздену жолдары әрқилы және әртүрлі болады [310, 142-143].

Академиялық грамматикада берілетін анықтама төмендегіше: «*адамның тек ақыл-ой қабілетінің жәрдемі нәтижесінде ғана түсінілетін түрлі ұғымдар: наным, шама-шарық, ынсап, парық, мінез-құлық, кәде-жора, ырым, пікір, тентектік, ұялашақтық, ырыс*» [311, 442].

Татар тілінің зерттеушісі А.Р. Нургатинаның пікіріне қарағанда, түркологияда дерексіз зат есімдер жеке зерттелмеген, тек белгілі түркологтар М. Казем-Бектің, Н.К. Дмитриевтің, Н.А. Басқақовтың, А.Н. Кононовтың, М.З. Закиевтің, Ф.А. Ганиевтің, Д.Г. Тумашеваның, Г. Ибрагимовтың, М. Курбанғалиеваның және басқа да бірқатар зерттеушілердің еңбегінде морфология аспектісінде қарастырылған [312, 6]. Шындығында да дерексіз зат есімдерді талдауға

арналған жеке-дара, кешенді, жүйелі түрдегі зерттеулер жоқтың қасы, оның үстіне түркологтардың барлығы дерлік әсіресе сөзтудырушы *-лық* (варианттарымен қоса алғанда) форманты арқылы жасалатын дерексіз зат есімдерге кеңінен тоқталғанын байқаймыз. Профессор Қ. Жұбановтың зерттеулерінде де бұл мәселе сөзтудырушы қосымшалардың өнімділігіне қатысты талданады [132, 379]. Арнаулы зерттеу жүргізген А.Р. Нургатинаның өзі де сөзтудыру форманттары арқылы жасалған дерексіздерді ғана негізгі нысана етіп алған.

Орыс тіл білімінде еңбектер бар. Бірақ оларда да дерексіз зат есімдердің туынды тұлғалары (қосымшалар арқылы жасалғандары) көбірек айтылады. Осыған байланысты шетел ғалымы Е.С.Кубрякованың «күрделі категория ретінде сөз таптарын ажыратудың, топтастырудың принциптерін белгілеудің байырғыдан келе жатқан дәстүрі бар десек те, бұл саланың әлі түйінді тұстары бар» деуі әбден орынды [313, 7]. Шындығында да, қазақ тілтанымында сөз таптарына, оның ішінде зат есімдерге қатысты тұжырымдардың берік, ғылыми іргетасы қаланған, дей тұрсақ та тіл ғылымындағы жаңа бағыттардың әдіснамаларына сүйене отырып, әлі де болса жетілдіретін, нақтылайтын тұстары бар екендігі анық.

Зерттеулердің негізгілеріне шолу нысанаға алынған проблеманың аса күрделі екендігін көрсетеді. Осы себепті Ұлттық корпуста дерексіз зат есімдердің лингвистикалық паспортын нақтылау мақсатында алдымен бірнеше мәселе баяндалады:

- дерексіз зат есімдерді тану.
- дерексіз зат есімдердің топтастырылуы.

Дерексіз зат есімдерді тану. Дерексіздік – философия, психология, логика ғылымдарының негізгі аспектісі. Осы салалардың дерексіздікті ажырату, тану бойынша негізгі әдіснамаларымен танысу негізінде біз үшін маңызды болғаны – дерексіздік деректілікке қарама-қарсы қойылып анықталады. Аса күрделі құбылысқа қатысты барлық пікірлерді ықшамдап көрсеткенде, ғалымдар «қарапайым сезімдік абстракция, қорытындылайтын абстракция, оқшаулайтын абстракция, мазмұндық абстракция деп аталатын түрлі-түрлі бөліністерді атайды. Шетелдік зерттеушілердің еңбектерінде сондай-ақ абстрактілі және нақты емес зат есімдер (отвлеченные существительные) деген де таратулар кездеседі, яғни честь, долг, отвага – абстрактілі түсініктер; ал бег, учение, выступление дегендер – нақты емес зат есімдер. Осының өзінен-ақ дерексізденудің де, дерексіздіктің де аса күрделі құбылыс екендігін аңғаруға болады. Логика, философия, психология салалары бойынша «дерексіздік дегеніміз не» дегенге анықтама да осы бөліністерге негізделеді. Сөйтіп, «дерексіздену – нысанның (заттың немесе құбылыстың) мәнді, заңдылық болып саналатын белгілерін айрықшалап көрсету мақсатында жүретін ойлау операциясы, дерексізденудің нәтижесі – дерексіз ұғымдар, мысалы, түс, қисық, салмақ, сұлулық» (Новоселов, 2010: 106).

Алдымен айтарымыз – А. Байтұрсынұлы мен А. Ысқақовтың анықтамалары «Қазақ грамматикасындағы» «адамның тек ақыл-ой қабілетінің жәрдемі нәтижесінде ғана түсінілетін түрлі ұғымдар» деген түсіндірмеге қарағанда, барынша нақты әрі өзара мазмұндас. Әрине, А. Ысқақовтың анықтамасы шек-

теулі; көзбен көріп, қолмен ұстауға болмайтын, ойлау, топшылау қабілетінен пайда болғандар ғана – дерексіздер. Оның үстіне осы еңбектердегі мәселенің анығырақ түсіндіруге болатын тілдік фактілер де шектеулі, өзара ұқсас. Бұлай деуімізге себеп – Ұлттық корпуста дерексіз зат есімдердің лексика-семантикалық белгіленімін анықтау мақсатында жинақталған материалдар мазмұн тұрпаты жағынан сан түрлі. Бұл ретте дерексіздік туындататын *-лық* (*варианттарымен қоса алғанда*) форманты арқылы жасалатын дерексіз зат есімдерге тоқталмауға да болады. Өйткені *-лық* (*варианттарымен қоса алғанда*) қосымшасының грамматикалық мағынасы, функционалдық қызметі, жоғарыда атап өткеніміздей, түркологияда да, қазақ тіл білімінде де әбден қарастырылған. Кезінде А. Байтұрсынұлының өзі осы формантты басқа нәрселерге өлшеу болатындығын білдіретін мағынасын ажыратқан, сөйтіп, «ай» деген сөзге *-лық* жұрнағын жалғап «айлық» дейміз, айдың өлшеу болатындығын көрсету үшін» деп жазған болатын [309, 212-213]. Кейінгі кезеңдердегі еңбектерде сөз тудыру қабілеті аса құнарлы қосымша болып саналатын «*-лық* жұрнағының (варианттарымен бірге) деректі және дерексіз заттық ұғым атауларын өнімді жасайтындығы тұжырымдалған: белдік, мойындық, кездік, орындық; тұздық, уыздық; патшалық, ұлықтық, билік; батырлық, құпиялық, қастықтық; қажығандық, түңілгендік» [310, 34-35]; [314, 211-218]. Дәлелдеулер барынша анық болғандықтан, абыройлылық, абыройсыздық, абайсыздық, абзалдық, абыржушылық, адалдық, ажарлылық, әдемілік, айбындылық, ақжарқындық, әбестік, әбігерленушілік, әсірешілдік, баққұмарлық, бақыттылық дегендердің дерексіз зат есім тобына енетіндігі шәк келтірмейді. Демек, осы типтес сөздердің сөзжасам құрылымы олардың морфо-семантикалық белгіленімін, «дерексіз» екендігін анықтауда қиындық жасамайды. Бұлар А. Байтұрсынұлынан бастап беделді зерттеушілердің еңбектеріндегі «көзге көрінбейтіндер, құлаққа естілмейтіндер, иісі сезілмейтіндер, денеге татылмайтындар, қолмен ұстауға, көзбен көруге болмайтындар, тек ақылмен ғана ажыратылатындар» делінген анықтамаға толық сәйкес келеді, сөйтіп, Ұлттық корпуста сенімді түрде «дерексіз» деген белгіленім көрсетіледі.

А. Байтұрсынұлы дауыс, дыбыс, ән, күй, иіс-қоңыс, дәм, ащылық, тұщылық, жылулық, суықтық дегендерді деректі зат есімдер тобына енгізген. Ал А. Ысқақовтың жіктелімінде және кейінгі еңбектердегі фактілер қатарында осы сөздер жоқ. Осы баяндалғандардан шығатын бастапқы екі қорытындыны көрсетуге болады:

– дерексіздікті тудыратын *-лық* (*варианттарымен бірге*) қосымшасы арқылы жасалған туынды сөздер – дерексіз (*әдемілік, адалдық, атаққұмарлық, әсірежағымпаздық, ақкөңілділік*);

– А. Байтұрсынұлының жіктелімі бойынша жер, су, ай, күн дегендер сияқты қолмен ұстауға, көзбен көруге болмайтын болса да, құлаққа естілетін, иісі сезілетін ұғымдар мен түсініктерді білдіретін сөздер – деректі [309, 199-200]; демек, біздің тарапымыздан жинақталған тілдік фактілер қатарындағы ажыл (желдің гуілі), ән, әуен, әуен-саз, арпыл, әупіл (иттің үруі), ән-күй (жиынтық атау), әнұран, айтыс (өнер үлгісінің үдерісі кезінде), абхаят (діни; өмір суы), ажар, ажар-көрік, ажар-сымбат, айғай, айқай, айғай-шу, айғай-сүрен, айқай-шу,

айғай-думан, айқай-қиқу аят (діни), ағуа (діни; уағыз) тәрізділер деректі зат есімдер қатарында;

– профессор А. Ысқақовтың анықтамасына сүйенсек, көзбен көріп, қолмен ұстауға болмайтын, ойлау, топшылау қабілетінен пайда болғандар, яғни эн, дауыс, айғай, ажыл (желдің гуілі) іспеттестер – дерексіздер;

– зерттеулерде нысананы анықтап, тануға септігін тигізетіндей тілдік фактілер аз.

Ұлттық корпуста лексика-семантикалық белгіленімі көрсетілетін тағы бір топ көлемді түбір және күрделі сөздер анықталды: абырой, абырой-атақ, абырой-бедел, ағым, адам-күн, адам-сағат, ажар, айбар, айбат, айбар-қоқай, айла, айнабесін, ақиқат, ақпар, ақшам, ауа, әбігер, әдеп, әдеп-иба, әдет, әдет-ғұрып, әдиса, әділет, әдіс, әдіснама, әдіс-тәсіл, әдістеме, әжір, әзіл, әзіл-әжуа, әлем, әлеует, әл-ауқат, баға, бағыт, бақ, бақ-талай, бақыт.

Жинақталған материал дерексіздер тобына жататын сөздердің құрылымы жағынан да күрделі екендігін көрсетеді: түбір сөздермен қатар (*атақ, әдеп, әділет*), түптөркіні араб-парсы тілдеріне барып тірелетін лексемалар (*әдиса, аят*), сөз біріктіру және сөз қосарлау арқылы жасалып, бірақ бір сыңары деректі болатын күрделі сөздер (адам-күн, адам-сағат), көне және жергілікті тіл ерекшеліктері (ай-зайра, әжір, әдірет), актуалды терминдер (әдістеме, әдіснама, абайтану, әуезовтану). Сонымен қатар дерексіз зат есімдердің ақиқат дүние болмысын, қайсыбір ұғымды атауы, қасиет-сапаны танытуы жағынан жинақтық мағынасы бар екендігі аңғарылады. Мәселен, айтыс – бір жағынан белгілі бір ғылым саласының термині, жанр түрі ретінде ұғымның атауы, сонымен қатар «айтысты тыңдап отырмын» деген сөйлемде «құлаққа естілетіндіктен» деректі мазмұнға ие деп тану керек болады.

Дерексіз зат есімдердің мағыналық ерекшеліктері. Дерексіздікті білдіретін сөздер – ең алдымен тілдің төлиесінің жеке өзінің пайымдауы мен ұлттық мәдениеттің, ұлттық философияның арасындағы тығыз сабақтастықты көрсететін, бірегей (уникалды) тілдік бірліктер. Екіншіден, материалдық заттарды атайтын деректі зат есімдерге қарағанда, ауызша сөзде және жазбасөзде де берілетін ақпараттың логикалығын, әсерлі болуын, адресатқа ықпал ету күшін арттыратын, мәтіннің басты құралы болып табылатын актуалды қолданыстағы сөздер тобы деп бағалауға болады. Өйткені осы сөздер – тобы адамның ойлау қабілетінің жемісі. Ал тілдік қарым-қатынасқа түсу когнитив санадағы үдерістерге байланысты екендігі белгілі. Бұл жөнінде Ұлт ұстазы – А. Байтұрсынұлы былай дейді: «Ойлау екі түрлі: адам ойлағанда иә нәрсеге тиісінше ойлайды, иә көңілінің түйісінше ойлайды. Нәрсе түрлі сипатты болады. Ол сипаттарының біреулері бойына біткен сипат болса, екіншілері адамның қосқан, таңған сипаттары болады [315, 394].

Дерексіз зат есімдер мағынасы жағынан ақиқат дүниені, қоршаған ортаның, материалдық заттардың күллі сипатынан, оларға адамның қарым-қатысынан туындайтын құндылықтарды, қасиет пен сапаны ұлттық философия деңгейінде, ұлттық мәдениет, ұлттық психология, әлеуметтік көзқарас тұрғысынан бағалайды, сөйтіп, ақиқат дүниенің барлық ерекшеліктерін сипаттап береді.

Мәселен, абырой, әдеп, әдет, иба, әжір (көне) ұлттың рухани мәдениетінен бастау алатын құндылықтардың қатарына жатады. Абзалдық, әдептілік, әдемілік, айбындылық, ақжарқындық, адалдық, ажарлылық, аккөңілділік, айтыскерлік, айтысқыштық әлеуметтік баға берудің жағымды жағы болса, абыройсыздық, абайсыздық, әбестік, әсірешілдік, айлакерлік, әсірежағымпаздық, баққұмарлық бағалаудың жағымсыз жағы болып табылады, ал абыржушылық, әбігерленушілік тәрізділер әрекет нәтижесі; *әлеует*, *әл-ауқат* – сапаның деңгей көрсеткіштері. Осылайша дерексіздердің ең негізгі интеграл мағыналарын құндылық, бағалау, сапаның деңгей көрсеткіштері деп анықтауға болады.

Дерексіз зат есімдердің лексикалық белгіленімінің үлгілері. Ұлттық корпуста, тұтастай алғанда, сөзді лексикалық семантикалық белгілеу, метасуреттеу, метамәлімет беру өте маңызды әрі күрделі. Өйткені әрбір төл сөз – мәдени-ұлттық код. Қазақ халқының мәдени-ұлттық коды – ұлт болып қалыптасқан кезеңнен бастап халықтың атау беру, образ жасау, баға беру сияқты стереотиптері сөздің бойына жинақталады. Осындай күллі ақпаратты жеткізудің, сақтап қалудың маңызы зор. Профессор Э.Д. Сүлейменованың көрсетуінше, жалпы Ұлттық корпустың жасалуы «тілді зерттеудің күшті қаруы ғана емес, сондай-ақ жаңа идеология, ол тілдің нақты қолданылуын зерттеуге, корпустық тілдік жоспарлауды жүзеге асыруға ғана емес, сонымен бірге қазақ әдеби тілін нығайтуға көмектеседі [316, 79]. Әрине, бұл ретте түрлі лексикографиялық еңбектер, әсіресе қазақ сөзі түгенделген Қазақ әдеби тілінің түсіндірме сөздігі айрықша орын алады. Дегенмен заманауи технология ақпарат алу мен ақпарат берудің күшті құралы болып отырған қазіргі кезеңде Ұлттық корпус – «бұл тіл мамандары мен лексикографтардың көптеген ұрпақтарының еңбегімен жасалған сөздіктер мен грамматикаларды қоса алғандағы мүмкіндіктерінен асып түсетін тіл туралы тамаша, бірегей, жаңаша анықтама» [316, 78]. Әсіресе қазіргі жаһандану заманының негізгі қарым-қатынас құралы болып келе жатқан өзге тегеурінді тілдердің қатарында ұлт тілінің өміршеңдігін сақтап қалу, толық әлеуетін таныту ісінде әрбір сөздің семантикалық сипатын белгілеу өте маңызды. Лексика-семантикалық белгіленім әрбір сөздің мағынасын ашу арқылы тіл мен ұлт тарихының, тіл мен ұлт мәдениетінің, тіл байлығы мен сөздің функционалды қызметінің арасындағы байланысты танытады, тілді қолданушының тіл байлығы жөніндегі ақпаратын кеңітеді. Бұл жөнінде ғалым А. Фазылжан нақ тұжырым жасайды: «раскрытие этих культурных кодов, сокрытых в этномаркированных (лингвокультурологических, неэквивалентных, лингвокогнитологических и др.) языковых единицах имеет не только чисто научную значимость, а также высокое практическое значение для казахской этноязыковой общности в целях цементирования его идентичности, а также сохранения духовной безопасности и укрепления иммунитета национальной культуры и духа против вестернизации, девальвации духовных ценностей. И именно корпус как инновационно-информационная открытая система, доступная в онлайн-режиме, предоставляющая пользователю массив казахских аннотированных текстов, был бы мощным и эффективным проводником таких – лингвокультурологических знаний, сокрытых в культурной семантике языковых единиц. Раскрытие этих культурных кодов, сокрытых в этномарки-

рованных (лингвокультурологических, неэквивалентных, лингвокогнитологических и др.) языковых единицах имеет не только чисто научную значимость, а также высокое практическое значение для казахской этноязыковой общности в целях цементирования его идентичности, а также сохранения духовной безопасности и укрепления иммунитета национальной культуры и духа против вестернизации, девальвации духовных ценностей. И именно корпус как инновационно-информационная открытая система, доступная в онлайн-режиме, предоставляющая пользователю массив казахских аннотированных текстов, был бы мощным и эффективным проводником таких – лингвокультурологических знаний, сокрытых в культурной семантике языковых единиц» [14, 103].

Демек, Ұлттық корпус – тіл дамуының барлық кезеңдеріндегі тіл туралы мәлімет жинақталатын, нақты анықтамалар (белгіленімдер) арқылы реттелген, теңгерімге түскен тілдік бірліктердің ерекше қазына-қоры.

Ұлттық корпусстың әлемдік тәжірибесіне қарағанда, белгіленім – сөздің мәнмәтіннен тыс та, дискурс ішіндегі де барлық табиғатын танып-білуге айрықша көмек көрсететін, корпусстық лингвистика саласының негізгі түсінігі. Метабелгіленім мәтінге қатысты бүкіл дереккөзді камтиды, мәтіннен тыс берілетін метасуреттеулер сөздің құрылымдық сипаттамасын ашып береді. Бұл атрибуттар, сөз жоқ, маңызды.

Ұлттық корпуста дерексіз зат есімдерді белгілеу күрделі. Біріншіден, лексикалық белгіленім тілді пайдаланушыларға сөздің негізгі мағынасынан толық хабардар етуі керек. Бұл жағынан алғанда, түсіндірме сөздіктің мүмкіндіктері кең. Ал Ұлттық корпусстың құрылымдық талаптары бойынша белгіленім барынша қысқа, әрі нақты болуы керек. Қазақ корпусстық лингвистикасының негізін қалаған профессор А. Жұбанов пен профессор А. Жанабекова белгі-код терминін қолданған [12]. Жалпы, код термині информатика саласында ақпарат беретін сан алуан символдардың шартты таңбасы деп түсіндірілетіні белгілі. Демек, лексикалық-тақырыптық белгіленімді – белгі-кодты беруде қазақ сөзінің күллі мағынасын жиынтықтап беретін сөз алынуы тиіс. Сонымен қатар бір тақырыптық топқа енетін тілдік бірліктердің мәліметі әртүрлі болуы да мүмкін. Сөзтудырушы *-лық (варианттарымен бірге)* форманты арқылы жасалған дерексіз зат есімдерге ықшам белгіленім беру мұндайда жеңіл: өйткені осы қосымшаның өзі дерексіз зат есім жасайды; түбір сөздің өз мағынасы да белгіленімді анықтап береді. Мысалы, адалдық, ақкөңілдік, әдемілік, ажарлылық, әдептілік, әдепсіздік сөздерін қасиет/сапа; жағымды / жағымсыз деген белгіленім арқылы шектеуге болады. Дегенмен, екіншіден, бірыңғай мағыналас (интеграл мағыналас) дерексіз ұғымдардың атауларының өзіне әртүрлі белгіленім таңдау қажет болады. Өйткені дерексіз зат есімдердің ауқымды тобы ұлттық мәдениетке тән рухани құндылықты, қасиет пен сапаны, психологиялық, эмоциялық, экспрессивтік, әлеуметтік жай-күйді, мағына арқылы танылатын деңгей көрсеткіштерін, әртүрлі бағалауды білдіреді. Шын мәнінде, әсіресе дерексіз зат есімдердің белгіленімін анықтауда дереккөздердегі (түсіндірме және нормативті сөздіктердегі, жазба нұсқалардағы) қолданысын зер сала отырып, стильдік тарамдалуына, ұғымдық аппаратына шығармашылық ізденіспен қараудың маңызы зор. А. Жұбанов қазақ тілі мәтіндерін туында-

ту моделінде қолданылатын зат есімнің семантикалық топтарын ақиқат дүниенің сегменттеріне ажыратады. Олардың қатарында ғалымдар көрсеткен жалпыфилософиялық ұғымдар, психологияның негізгі ұғымдары, адам және оның әлемі, адамдардың арақатынасына қатысты ұғымдар [2014, 149-150]] деген семантикалық топтар дерексіз зат есімдерге тікелей қатысты.

Әділет сөзі – ақиқат дүниенің сегменті. Бұл сөзге берілген анықтама төмендегіше: «әділет [ар. عدالة] зат. қоғ.-әлеум. Қоғам өміріндегі адамдық қарым-қатынас, заң мен тәртіп, құқықтар мен міндеттер, қоғамдық, табиғи мүліктің бөлінісі сияқты т.б. әлеуметтік маңызды тіректердің тұрақты қалыптасқан ахлақ нормаларына сәйкестігі» [317]. Сөз жоқ, түсіндірме сөздің мағынасын анық, нақты көрсеткен. Алайда белгіленім үшін «ақиқатқа сәйкестік» анықтамасын (барлық түсіндірмені топтастыратын интеграл мағына) ғана беруге болады.

Әдеп сөзі – стилдік тарамдалу жағынан алғанда, әрі қоғамдық-саяси, әрі ұлттық психологиялық, ұлттық философиялық, әрі әлеуметтік термин; адам және оның әлемі, адамдардың арақатынасына қатысты іргелі ұғымдардың бірі. «Белгілері бір халықтың әдет-ғұрпында тұрақтанып қалыптасқан этикалық тәртіп» [318]. Жалпы қоғамға ортақ әдеп нормалары бар, жеке адамның өзінің де әдеп нормасы деп танитын қағидалары болады. Әдеп ұғымы, тұтастай алғанда, адамдардың мінез-құлқын реттеуші ережелердің жиынтығы. Ал әдеп-иба, Түсіндірме сөздіктегі дереккөздерге қарағанда, көбінесе ортаға түскен жеке адамның тарапынан танылатын норма. Есенейге: – Сен ғой, өз үйіңде Есенейсің! Күйеу дегеннің әдеп-ибасынан хабарың жоқ, – деді (Ғ. Мүсірепов) [317] деген сөйлемде өзара байланысқан күрделі сөздегі мағыналардың ұйымдасуы жеке адамның ортадағы статусына қатысты норманы анықтайды. Бұл мағынаға иба сөзі тірек болады; «Иба – адамның мінезінің әдеп сақтап, ізет етуін, ұялу сезімін білдіретін ұнамды әрі жағымды қасиетін білдіретін ерекшелігі [319, 37]. Яғни әдеп нормалар жиынтығы болса, иба – жеке адамның ерекшелігі; жалпыға ортақ ізгілік нормасы, демек, екі лексеманың мағынасы ұлттық мәдениетке, ұлттық этикаға негізделген. Сондықтан өзара интеграл мағыналас осы анықтамаларды негізге ала отырып, корпустың әдеп, әдеп-иба сөздері үшін Ұлттық корпуста ортақ ықшам түрдегі, нақты лексикалық анықтаманы (құндылық) шығаруға болады.

Дерексіздікті білдіретін сөздердің белгіленімі негізінен: а) қасиет, ә) сапа; б) құндылық; в) жағымды / жағымсыз; г) деңгей көрсеткіштері арқылы анықталады. Дегенмен тілдің қазіргі төлиесінің тіл байлығы жөніндегі ақпарат өрісін кеңейту, тіл байлығын таныту мақсатында кейбір көне, сирек қолданыстағы сөздердің, түптөркіні кірме сөз болып келетіндердің, әдеби тілге ене бастаған, көркем туындыларда қолданылатын жергілікті тіл ерекшеліктерінің ішінара синонимін, кей реттерде барынша ықшам түсіндірмесін беру қажет. Мәселен, АЙ-ЗАЙРА – төл/көне; мифологиялық термин; фантастикалық ұғым; мағынаны дәл көрсету үшін микротобы ретінде періште сөзі беріледі. ӘЖІР төл / көне / жерг.; стилдік таралымы әлеуметтік деп анықталады; рухани мәдениет, құрмет сөзімен мағыналас; алайда негізгі мағынасын ашатын лексема – разылық. Сол сияқты ӘБЖІЛДІК сөзі де жағымды/ширақтық деген белгіленімді, яғни жағымды қасиетті танытатын актуалды, концептуалданған синонимді (ширақтық) беруді қажет етеді. Бұған себеп – көркем шығармаларда және көсемсөзде қолданысқа түсе

бастағанына қарағанда, әбжілдік жергілікті тіл ерекшеліктері қорынан әдеби нормаға ене бастаған сөздердің қатарында деп танымыз.

Талқылаудың нәтижесінде Ұлттық корпуста «көзбен көруге, қолмен ұстауға болмайтын» ұғымдарды атайтын дерексіз зат есімдер топтастырылған үлгі ұсынылады.

Кесте 23

Дерексіз зат есімдердің лексикалық метабелгіленімінің көрінісі

Дерексіз зат есім	Лексикалық қабаты	Терминдік қолданысы	Тақырыптық-семантикалық тобы	
			Макротоп	Микротоп
абырой	төл	әлеум.	құндылық	жағымды бедел
алабауыр	төл	түйешар. вет.	мал ауруы	індет
ай-зайра	төл көне	миф.	фантастика	періште
айла-қулық	төл	әлеум. психол.	қасиет әрекет	жағымсыз
аңызқ	төл	табиғат	құбылыс	аптап жел
әдемілік	төл	әлеум. психол.	құндылық	сапа деңгей көрсеткіші
әдеп	төл	мәд. этика	рухани мәдениет	құндылық
әдепсіздік	төл	қоғ.-саяси психол. әлеум.	қасиет	жағымсыз
әдиса	төл	психол. поэт.	сапа	құпия сыр
әділет		әлеум. психол.	құндылық	ақиқатқа сәйкестік
әдіснама	төл	пед. оқу-ағарту ғылым	жүйе	әдіс-тәсілдер жүйесі
әлеует	төл	қоғ.-саяси әлеум.	сапа	деңгей көрсеткіші
әл-ауқат	төл	қоғ.-саяси әлеум.	сапа	деңгей көрсеткіші
ән	төл	мәд. муз.	рухани мәдениет	жанр
әуезовтану	төл	әдеб.	ғылым	сала
байғазы	төл	этн. мәд.	ұлттық мәдениет дәстүр	жоралғы
бақыт	төл	филос.	жай-күй сапа	жағымды деңгей көрсеткіші
баққұмарлық	төл	психол.	қасиет	жағымсыз

Адамның ойсанасында қоршаған орта жөнінде ақпарат жинақталады, сақталады және өңделеді, ал өңделу процесі тіл қолданысқа түскен кезде жүріп отырады, адамның ойсанасындағы тілдік бірліктің семемасы немесе семасы да қайта құрылады. Ақпаратты логикалық тұрғыдан анық, адрестқа ықпалды жеткізу мүддесінде ойсанадағы бар сөз «прототип» ретінде алынады; бұл жаңа сөз ойлап тапқаннан гөрі, басқа тілдің сөзін қолданғаннан гөрі әсер туғызу, ықпал ету үшін ыңғайлы болады. Осындай күрделі құбылыс әсіресе дерексіз зат есімдерге қатысты. Сондықтан деректілікті білдіретін зат есімдерге қарағанда, дерексіз ұғымдарды метасуреттеу философия, психология, логика, психоллингвистика, тілтаным салаларының әдіснамалары арқылы зерттеу елегінен өткізуді қажет етеді. Сонымен қатар әлемдік лингвистикадағы, қазақ тіл біліміндегі зерттеулер дерексіздікті білдіретін тілдік бірліктердің әлі де болса тереңірек қарастырылатын мәселелері бар екендігін көрсетті. Мәселен, «ойлау, топшылау қабілеті арқылы танылатындар», «адамның тек ақыл-ой қабілетінің жәрдемі нәтижесінде ғана түсінілетін түрлі ұғымдар», мазмұндық абстракциялану, сезім абстракциясы, қорытындылаушы абстракциялану тәрізді анықтама-бөліністердің нақты түйіні, эмпирикалық дәлелденуі әлі де айқын емес. Бұл ретте дерексіз ұғымдарды білдіретін түрлі лексемалардың стилдік қолданысы, макротоптық және микротоптық мағыналарының жинақталып, Ұлттық корпуста берілуі, біріншіден, тіл мен мәдениеттің арасындағы сабақтастық, сөз мағынасының күрделілігі туралы мәліметті қамтып, тілді қолданушының лингвистикалық өрісін кеңейтеді, қазақ тілінің байлығына тұшындырады. Сонымен қатар Ұлттық корпуста әр сөздің лексика-семантикалық мағынасының ашылуы әсіресе білім беру жүйесінде, лексика-грамматикалық минимум, басқа да лингводидактикалық база жасауда айрықша маңызды дерекқор болып саналады.

11.2 Зат есімдердің лексика-семантикалық белгіленімін ұлттық корпуста рәсімдеу жолдары

Лексика-семантикалық белгіленім – Корпусқа енгізілген мәтін бойындағы сөздердің лексика-семантикалық мәліметінен хабар беретін таңбалық жүйе. Лексика-семантикалық белгіленім не үшін керек? Сөйлемдегі сөздің мағынасын тану үшін қажет. Мысалы: Сатыға құрт жайылған екен деген сөйлемдегі «саты» сөзін тану үшін үстінен басқанда лексика-семантикалық белгіленімі былайша рәсімделіп шығады:

Саты 1 – еңбек құралы

Саты 2 – үйдің баспалдағы

Саты 3 – өре

Қалмақ тілінің ұлттық корпусында лексика-семантикалық белгіленім «семантикалық аннотация» деп аталады. Орыс тілінің корпусынан сәл өзгеше. Бірақ мақсат бір – қолданыстағы мағынасын анықтау. Мысалы, қалмақ корпусында «хуучн» ескі мағынасындағы сөз тек жансыз зат есімдермен ғана тіркеседі де, «квгшн» ескі мағынасындағы сөз (кәрі) тек жанды зат есімдермен тіркеседі екен. Осы мағыналық белгісін анықтау корпуста рәсімделеді.

Қазіргі уақытта Корпус мәтіндерінің ішінара семантикалық жіктелуін табу лексикалық және семантикалық ерекшеліктерге негізделген іздеу жүйесі арқылы жүзеге асырылады. Лексика-семантикалық белгіленім арқылы мәтіндегі сөздердің бір немесе бірнеше мағыналық және туынды белгілер тануға мүмкіндік бар. Семантикалық белгіленім Корпус қорына жүйелі түрде белгілі бір реттілікпен енгізіледі. Алдымен барлық сөз таптарына ортақ тақырыптық топтар дайындалады. Бұл тақырыптық топтарға әрбір сөздің тура лексикалық мағыналары алынады. Ал ауыспалы мағынадағы семантикалық белгіленімі әрбір сөз табының жеке ұяшығында беріледі. Мәселен, алтын – металл атауы ретінде жалпы тақырыптық топта зат есім ретінде тіркеледі. Дегенмен, алтын – бағалы, құнды деген сындық ауыспалы мағынасы ретінде сын есімнің ішкі семантикалық белгіленімінде де тіркеледі. Сол сияқты «жанат» – жануар, оның ішінде тышқан түрі болғандықтан, тақырыптық топ қатарында жануарлар бөлігінде семантикалық жіктеліміне талданады. Содан кейін барып, сын есімнің классификациясында жанат сөзінің ауыспалы мағынасы – «ең бағалы, құнды» мәнінің белгіленімдері анықталады. Корпус базасындағы семантикалық белгіленім әрбір сөздің 5-тен 50-ге дейінгі белгісін тануға мүмкіндік бере алады. Сондықтан кейбір сөз таптарының семантикалық жіктелімі макро және микро топтар іріктелген. Ал мәтін ішіндегі сөздің барлық белгіленімдері анықталғанда, сол мәтіндегі мағынасын тану оқырманның еншісінде болмақ. Егер сол мәтіндегі сөздің мағынасымен дәл тану сөздердің көпмағыналы базасын енгізгенде жүзеге асады.

Енді корпус мәтіндеріне семантикалық белгіленім рәсімдеудің әлемдік тәжірибелеріне тоқталайық. Орыс тілінің ұлттық корпусында семантикалық белгіленім «Лексикограф» жүйесіне негізделеді. Бір сөз бірнеше тапқа жататын фасетті жік-жікке бөледі. Бірінші кезеңде сөздіктегі белгілер бойынша ізденіс жүргізіледі. Корпустың семантикалық сөздігіне сәйкес Semmarkup бағдарламасының (авторы А.Е.Поляков) көмегімен автоматты түрде жүзеге асырылады. Семантикалық белгіленімдерді мәтінде қолмен өңдеу қажырлы еңбекті қажет ететіндіктен, Орыс тілінің ұлттық корпусындағы семантикалық омонимия жойылмайды, көп мағыналы сөздерге семантикалық белгілердің бірнеше баламалы жиынтығы тағайындалады.

Орыс тілінің ұлттық корпусында мәтіндегі әрбір сөзге берілетін лексика-семантикалық ақпарат метаразметканың үш тобынан тұрады:

1. Категория;
2. Лексика-семантикалық сипаттамалар (тақырыптық топ, себеп-салдарлық белгілер, бағалаулар);
3. Туынды (сөзжасамдық) белгілер.

Лексика-семантикалық ақпарат әртүрлі құрылымға ие. Мысалы, зат есімнің семантикалық белгіленімі – субъектілік, субъектілік емес және жалқы/жалпы, жанды/жансыз топтарына басымдық бергені аңғарылады.

Ал тақырыптық топтарға жіктегенде семантикалық белгіленімдердің өрістері кеңейеді. Мысалы:

1. таксономия (лексеманың тақырыптық класы) – зат есім, сын есім, етістік және үстеу үшін;

2. мереология («бөлік – бүтін», «элемент – жиынтық» қатынасының көрсеткіші) – объективті және объективті емес атаулар үшін;

3. топология – пән атаулары үшін;

4. себептілік – етістіктерге;

5. ресми мәртебе – етістіктер үшін;

6. бағалау – пәндік және пәндік емес атауларға, сын есімдер мен үстеулерге.

Бұнымен қоса, сөзжасамдық сипаттамалардың бірнеше түрлері семантикалық белгіленім ретінде анықталған. Мысалы:

– морфо-семантикалық туынды белгілер (мысалы, «кішірейткіш», «карикативтік», «семельфактивті»);

– сөз тудырушы категориясы (субстантивтену, т.б.);

– сөзжасамның морфологиялық түрі (күрделі сөз, біріккен сөз, тіркесті сөз).

Ағылшын және орыс тілінің ұлттық корпусындағы лексико-семантикалық белгілер қысқартылған таңбалар («тегтер») жүйесімен рәсімделген. Корпус пайдаланушыларының кең халықаралық аудиториясын ескере отырып әзірлегені байқалады. Орыс тілінің Ұлттық корпусындағы зат есімнің лексико-семантикалық белгіленімі былайша берілген:

Разрядтар

г:concr	деректі атау (үстел, қалам)
г:abstr	дерексіз атау (жарық, уақыт)
г:propn	адам есімі (Иван, Эйнштейн, Петроград)

Заттық атаулар

Таксономия

t:hum	адамзат (адам, мұғалім)
t:hum:etn	этнонимдер (эфиоп, итальян)
t:hum:kin	туыстық атаулар (аға, әже)
t:hum:supernat	мифтік атаулар (русалка, инопланетянин)
t:animal	жануар атаулары (сиыр, жираф, сауысқан, кесіртке, құмырсқа)
t:plant	өсімдік (қайың, раушан, шөп)
t:stuff	бұйым атаулары (жиһаз, әшекей)
t:space	кеңістік атаулары (аспан, жайлау, орман, ғарыш)
t:constr	ғимарат, мекеме атаулары (үй, көпір)
t:tool	жабдық атаулары (сіріңке, түйме, ине)
t:tool:instr	құрал атаулары (балға, қалам)
t:tool:device	тетіктер мен аспап атаулары (телефон, қысымөлшеуіш)
t:tool:transp	көлік атаулары (автобус, пойыз, шана)
t:tool:weapon	қару атаулары (қылыш, пистолет, шоқпар)
t:tool:mus	музыка аспаптары (рояль, скрипка, сыбызғы)
t:tool:furn	жиһаз атаулары (үстел, диван, шкаф)
t:tool:dish	ыдыс-аяқ атаулары (чашка, кастрюля, фляжка)
t:tool:cloth	киім мен аяқкиім атаулары (көйлек, шляпа, бәтеңке)
t:food	тағам мен сусын атаулары (пирог, ботқа, сүт)
t:text	мәтін атаулары (әңгіме, кітап, ертегі) [39].

Қазақ тілінде лексика-семантикалық өрісі аса кең сөз таптарының бірі – зат есім. Қазақ грамматикасында зат есімге берілген анықтамада: «Айналамызды қоршаған дүниедегі зат пен жан-жануар атаулыға, барша құбылыс пен уақиға-болмыс атаулыға есім, атауыш болып келетін сөздердің бәрі – зат есімдер. Лексикалық тұрғыдан алғанда зат есімдер өз алдына мағыналық дербестігі бар, заттық ұғымды білдіретін сөздер. Табиғат пен қоғамдық тіршіліктің қырсыры қаншалық алуан түрлі болса, олардың бас-басына дерлік атауыш ретінде жұмсалатын тілдегі сөздер де соншалық алуан түрлі. Зат есімдер тікелей сезім арқылы тануға болатын дүниедегі нақтылы нәрселерден бастап, адамның тек ақыл-ой қабілетінің жәрдемі нәтижесінде ғана түсінілетін түрлі абстракт ұғымдардың аттарын да білдіре алады», – делінген [311].

Осы еңбекте зат есімнің лексика-грамматикалық топтарын ішінара санамалап былай көрсетеді:

1. Бұйым мен нәрсе атаулары;
2. Тағам аттары;
3. Жан-жануарлар аттары;
4. Өсімдік атаулары;
5. Табиғат денелері мен құбылыстарының атаулары;
6. Адамның кәсібін, тегін, мамандығын, қабілетін, қоныс-тұрағын білдіретін атаулар (мұғалім, алыпсатар, найман, қазақ, саятшы);
7. Тұрмыс-салтқа, әдет-ғұрыпқа байланысты айтылатын сөздер (айтыс, жар-жар, қыз куу);
8. Адамзат және ғаламзат есімдер;
9. Жалпы және жалқы есімдер;
10. Көптік мәнді есімдер;
11. Реңдік мәнді есімдер;
12. Эмоция-экспрессия реңі бар есімдер;
13. Көмекші есімдер.

Зат есімнің лексика-семантикалық топтарына қатысты бірнеше ғалым жіктелім жасаған болатын. Солардың ішінде ҚТҰК-дағы зат есімдердің лексика-семантикалық белгіленімін әзірлеуде ғалымдар Асқар Жұбановтың, Мейірбек Оразовтың, Бағдан Момынованың зерттеулеріне тоқталдық.

Кесте 24

Асқар Жұбанов ұсынған жіктелім Қазақ тілі мәтіндерін туындату моделінде қолданылатын зат есімнің семантикалық топтары [12].

Ақиқат дүниенің сегменттері	Семантикалық топтың атауы	Топ коды (шартты белгісі)
1	2	3
1. Ғылым	Жалпыфилософиялық ұғымдар	ЗТ101
	Жалпығылыми ұғымдар	ЗТ102
	Жалпымедициналық ұғымдар	ЗТ103

	Психологияның негізгі ұғымдары	ЗТ104
	Нейропсихологияның негізгі ұғымдары	ЗТ105
	Логиканың негізгі ұғымдары	ЗТ106
	Геометрияның негізгі ұғымдары	ЗТ107
	Негізгі географиялық ұғымдар	ЗТ108
	Тіл білімінің негізгі ұғымдары	ЗТ109
	Әдебиеттанудың негізгі ұғымдары	ЗТ110
	Бейнелеу өнерінің негізгі ұғымдары	ЗТ111
	Музыка өнерінің негізгі ұғымдары	ЗТ112
	Заң мен қылмыс салаларының негізі ұғымдары	ЗТ113
	Театрға қатысты негізгі ұғымдар	ЗТ114
	Авиацияға қатысты негізі ұғымдар	ЗТ115
	Құрылысқа қатысты ұғымдар	ЗТ116
	Механикалық нысанның жеке құрылысына байланысты ұғымдар	ЗТ117
	Ақпараттық ұғымдар	ЗТ118
	Жинақтық ұғымдар	ЗТ119
	Электроника мен электротехниканың негізгі ұғымдары	ЗТ120
	Әкімшілік ұғымдар	ЗТ121
	Мекемелердің, газеттердің, журналдардың және т.б. аттары	ЗТ122
	Абстракт есімдер	ЗТ123
	Баға	ЗТ124
2. Табиғат	Табиғат заттары	ЗТ201
	Табиғи құбылыстар	ЗТ202
	Табиғат бұйымдары – артефактар	ЗТ203
	Табиғаттың қалыптары	ЗТ204
	Заттар құрылысы мен оның бөліктерін сипаттайтын ұғымдар	ЗТ205
	Заттар, металдар	ЗТ206
	Табиғат заттарының әрекеттері	ЗТ207
3. Адам және оның әлемі	Адам және оның дене мүшелері	ЗТ301
	Адам беті мен басының бөліктері	ЗТ302
	Адамға тән күй мен сезім	ЗТ303
	Адамға тән әрекет	ЗТ304
	Адамдарды сипаттау үшін қолданатын ұғымдар	ЗТ305
	Адамның киімі мен аяқ киімдеріне қатысты заттар және оларға байланысты ұғымдар	ЗТ306

	Гигиеналық, әшекейлік және т.б. мақсатта адам қолданатын артефакт-заттар	ЗТ307
	Адам орын ауыстыруында қолданатын нысандар және оларға қатысты ұғымдар	ЗТ308
	Адам пайдаланатын азық-түлік заттары	ЗТ309
	Адамның тұрғын үйге қатысты ұғымдары	ЗТ310
	Адам өлімі мен жерлеуге қатысты ұғымдар	ЗТ311
	Адамдардың аты, әкесінің аты мен тегі	ЗТ312
	Адамның ұлты	ЗТ313
	Адам қолданатын ыдыс-аяқтар және оларға байланысты ұғымдар	ЗТ314
	Адамдардың демалыс кезінде, аңшылық, балық аулау кезінде пайдаланатын заттары және оларға қатысты ұғымдар	ЗТ315
	Мамандық, лауазым, ғылыми атақ, дәреже атаулары	ЗТ316
	Оқу мен еңбек ету және т.б. кезінде пайдаланатын артефакт-заттар	ЗТ317
	Адамдардың арақатынасына қатысты ұғымдар	ЗТ318
	Адамдар арасындағы туыскандық қатынастарға байланысты ұғымдар	ЗТ319
	Адам марапаттары	ЗТ320
4. Жануарлар әлемі	Ірі қара мал және олардың дене мүшелері	ЗТ401
	Ұсақ үй жануарлары және олардың дене мүшелері	ЗТ402
	Ірі жабайы аңдар және олардың дене мүшелері	ЗТ403
	Жәндіктер және олардың дене мүшелері	ЗТ404
	Құстар және олардың дене мүшелері	ЗТ405
	Жануарларға тән әрекеттер	ЗТ406
	Жануарлар, құстар мен жәндіктердің қорегіне қатысты заттар	ЗТ407
5. Өсімдіктер	Өсімдіктер және оның бөліктері	ЗТ501
	Ағаштар және оның бөліктері	ЗТ502
6. Уақыт	Уақытқа қатысты ұғымдар	ЗТ601
	Апта күндері	ЗТ602
7. Дін	Дінге қатысты ұғымдар	ЗТ701
8. Соғыс және оған қатысты ұғымдар	Соғыс техникасының заттары, қару-жарақ және оның бөліктері	ЗТ801
	Әскери мамандықтардың атаулары	ЗТ802
	Әскери атақтар	ЗТ803
	Соғыс кезіндегі көмек көрсетуге байланысты ұғымдар	ЗТ804

9. Оқу мен еңбек	Әскери бөлімшелердің аттары	ЗТ805
	Соғысқа байланысты ұғымдар	ЗТ806
	Оқыту мен еңбекке қатысты мекемелер	ЗТ901
	Оқыту мен еңбекке қатысты ұғымдар	ЗТ902
10. Меры	Ұзындық, салмақ, жылдамдық және т.б. өлшемдер	ЗТ911
11. География	Елді мекендердің аттары	ЗТ921
	Елдер, материктер, жер шары бөліктері, аралдар, таулардың және т.б. аттары	ЗТ922
	Өзендер, теңіздер, көлдер, бұғаздардың аттары	ЗТ923

Мейірбек Оразов ұсынған жіктелім

Дене мүшелері атауы, тамақ атауы, киім-кешек атауы, жануарлар атауы, құс атауы, ыдыс-аяқ атауы, жиһаз атауы, әшекей бұйым атауы, құрал-жабдық атауы, қару-жарақ атауы, өсімдік атауы, мезгіл атауы, өлшем атауы, кеңсе тауарлары атауы, туыстық атаулар, салт-дәстүр, адам атауы, қолөнер атауы, ойын атауы, кәсіп иесі атауы, аңшылық атауы, саятшылық атауы, киіз үй атауы, діни атаулар, музыкалық аспап атауы, аспан денелері атауы, табиғат құбылыстары атауы [320].

Бағдан Момынова ұсынған жіктелім

Адамға қатысты атаулар, сөзге қатысты атау, медицинаға байланысты атау, адамға, тіршілік иелеріне тән қозғалыс атау, үй және тұрмысқа қажетті бұйым атау, шаруашылық пен өнім атау, ғылымға қатысты атау, құжатқа қатысты атау, табиғатқа қатысты атау.

ҚТҰК-да зат есімнің лексика-семантикалық белгіленімін рәсімдеуде «Формадан – мағынаға қарай» қағидасы негізге алынды.

1. Реті
2. Морфологиялық белгіленімі – сөз табы – (зат е., сын е)
3. Сөзжасамдық белгіленімі
4. Морфо-семантикалық белгіленімі
5. Лексика-семантикалық белгіленімі

Кесте 25

Сөзжасамдық белгіленім

Сөзжасамдық белгіленім			
Сөздің морфемалық құрамы	Сөзжасамдық тәсілі	Тұлғасы	Құрамы
түбір; түбір+қосымша; түбір+түбір	синтетикалық/ аналитикалық/ лексика-семантикалық	негізгі/туынды	дара/ күрделі: қос сөз, біріккен, кіріккен

Морфо-семантикалық белгіленім

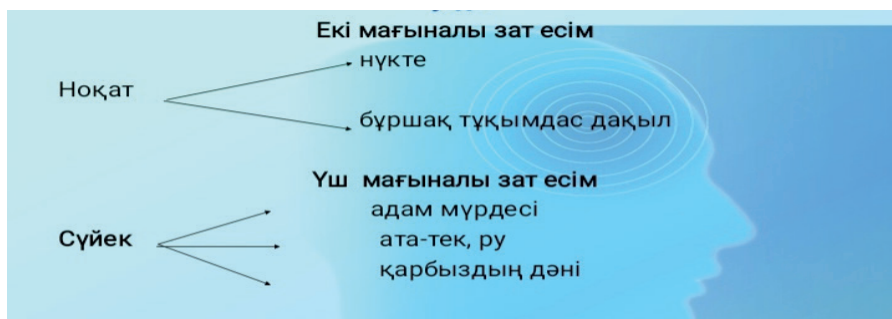
Морфо-семантикалық белгіленім		
Нақтылық/ абстрақтылық сипаты	Адамзаттық/ғаламзаттық сипаты	Жалпы/жалқылық сипаты
деректі/дерексіз	жанды/жансыз	жалпы/жалқы

Лексика-семантикалық белгіленім

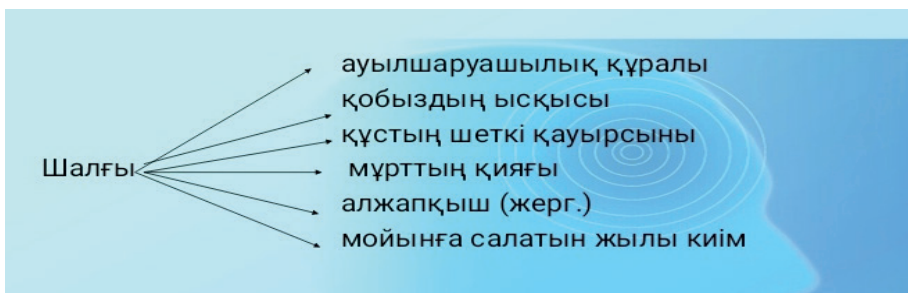
Лексика-семантикалық белгіленім				
Лексикалық қабаты	Стильдік ренкі	Терминдік қолданысы	Тақырыптық-семантикалық тобы	
			макротоп	микротоп
төл/кірме/шеттілдік/ жерг./ кәсіби/ тарихи/ көне/ жаңа/ кітаби	сөйлеу/ карапайым/экспр/ эмоц қарғыс/алғыс/ эвфемизм	спорт, өнер, муз., хим., мед., психол., авиаци., анат., аңшылық, археол., астр., ауылшар., әдеб., эск., биол., бот., геолог., геогр., геом., діни, заң, зерг., зоол. т.б. бот.		
Сарана - жергілікті			өсімдік	лалагүл

Қазақ тілінің семантикалық жүйесі (макро/микротоп бойынша) «жинақты/жіктелмелі» қағидасына негізделді. Мысалы, жинақты – ағаш. Жіктелімі – қарағай, емен, қайың. Жинақты – кәсіп иесі. Жіктелімі – мұғалім, дәрігер, аспаз, сатушы.

Зат есімдердің лексика-семантикалық белгіленімдерін анықтау негізінде көпмағыналы зат есімдер мен омонимдерді анықтауға мүмкіндік туады:



Сурет 29. Екі, үш мағыналы зат есім



Сурет 30. Алты мағыналы зат есім

Зат есім – жалпы грамматикалық мағынасы бойынша заттық ұғымды білдіретін сөздерден тұратын, олардың сандық, тәуелдік, септік, категориялары бар, баяндауыш қызметінде жақ категориясының көрсеткіштерімен түрленетін сөз табы. Заттық ұғым – зат есім сөздердің лексикалық мағыналарынан абстракцияланған зат есім сөздердің бәріне ортақ мағына. Тарихи тұрғыдан алғанда тіліміздегі зат есім, сын есім, үстеу деп аталатын қазіргі сөз таптарының әуелде бір негізден, біртұтас лексика-грамматикалық сөздер тобынан өрбіген құбылыс екенін айтады тіл зерттеуші білімпаздар.

11.3 Қазақ тілінің ұлттық корпусындағы етістіктердің лексика-семантикалық белгіленім әзірлемесі: әлемдік тәжірибе, классификациялау, корпуста белгілеу

XX ғасырдың басында қазақ тілінің оқулықтары жазылып, жеке сала қалыптаса бастады. Ал XX ғасырдың ортасына қарай әр саласы академиялық ғылыми зерттеу нысанына айналды. Одан бергі кезеңде тіл грамматикасы, лексикасы, фонетикасы, тарихы динамикалы зерттелді. Осы орайда Тәуелсіздік алғаннан кейінгі қазақ тіл білімінің жаңа кезеңінің ерекшелігі неде деген сұрақ туындайды. Бұл жолда өткенді саралау, сараптау және дамытудан басталып, әлемдік лингвистикадағы жаңа білімдер жүйесі меңгеріліп, әр тақырыпқа тереңірек бойлау қолға алынды. Яғни тіл дамуын қоғам дамуымен сабақтастыра отырып, жаңа заман жетістіктерін игеріп, меңгеріп отыру талап етіледі. Корпустық лингвистиканың қазақ тіл біліміндегі орнын осындай жаңару, жаңа технологияларды ұтымды пайдаланудағы басты жетістік ретінде атап айту қажет.

Инновациялық зерттеулер қазақ тіл білімінде Кеңес Одағы кезінде-ақ Қазақ ССР Ғылым Академиясының Тіл білімі институтында 1969 жылы өткен «Түркі тілдеріндегі статистикалық және ақпараттық зерттеу» атты Бүкілодақтық ғылыми жиыннан бастау алды. 1970 жылы математикалық статистика саласының маманы, белгілі математик-ғалым Қ. Бектаевтың жетекшілігімен Институтта «Тіл статистикасы және автоматтандыру» ғылыми тобы құрылады. Бұл ғылыми топ алғашқыда қазақ тілінің әртүрлі стильдері бойынша жиілік сөздіктер жасаумен айналысты. Мәселен, өткен ғасырдың 90-жылдары М.О.Әуезовтің 20 томдық шығармалар жинағы ЭЕМ жадына енгізіліп,

нәтижесінде 1995 жылы «М.Әуезовтің 20 томдық шығармалар текстерінің жиілік сөздіктері» жарық көрді. Одан беріде компьютерлік лингвистика, тілдік бірліктерді формальды модельдеу бағытында жұмыстар жүргізілді. Әлемдік тіл біліміндегі лингвистикалық ғылыми зерттеулерді цифрландыру, оңтайландыру, автоматтандыру үдерісі қазақ тіл біліміндегі қолданбалы лингвистика, әлемдік лингвистикадағы жаңа бағыт – корпустық лингвистика саласын зерттеу мәселесін көтерді.

Корпустық лингвистика әр ұлт тіліндегі мәтіндерді компьютерлік базаға жинақтап, оны программалық басқару жүйесіне салатын Ұлттық корпустарды құрастыру мен оны пайдалану мәселелерін зерттейді. Әлемдік тіл білімінде корпус жасау ісі 1960 жылдан Браун корпусынан басталады. Өзге тілдердің көбінде қазірде Ұлттық корпустар жасалған. Корпус жасау қазіргі ақпараттық технология заманында заманауи инновациялық құрал жасау деген сөз. Ұялы телефондарға орнатылған сансыз электрондық қызметтер орын алған бүгінгі жағдайда ғылыми зерттеу құралдарын автоматтандыру, жас зертеушілерге тиімді әрі қолжетімді ету, ғасырлар бойы жинақталған түрлі стильде жазылған жазба мәтіндерді, құнды шығармаларды т.б. рухани мұрамызды компьютерлік жадта сақтау – өзекті мәселе. ҚТҰК-ны жасауға қатысты жұмыстар қазіргі таңда Тіл білімі институтында қарқынды жүруде.

ҚТҰК-ның құрылымы, жүйесі, ондағы мәтіндер мен метабелгіленімдерге қатысты жазылған монографиялар, ғылыми мақалалар аз емес. Алайда әлі де тереңірек қарастыруды қажет ететін тұстары да бар. Соның бірі – корпустағы лексика-семантикалық белгіленімді жүзеге асыру мәселесі.

Корпустық лингвистиканы тереңірек түсіну үшін зерттеу барысында әлем тілдеріндегі корпустар назарға алынды, әр корпустың негізгі бағыттары, жұмыс істеу жүйесі анықталды. Аталған жұмыстар салыстыру, талдау негізінде жүзеге асырылды. ҚТҰК-ның құрылысы сипаттама, яғни синхрондық әдіс негізінде түсіндірілді. Етістіктердің лексика-семантикалық белгіленіміне қатысты теориялық ақпараттар мен практикалық талдаулар құрылымдық әдіс арқылы зерттелді. Лексика-семантикалық белгіленімнің қызметі мен құрылымы туралы тұжырым жасауда лингвистикалық әдістер басшылыққа алынып, корпустағы бейнесі мысалдар арқылы көрсетілді. Қазақ тілінің бай лексикалық қоры, әсіресе қазақ тіліндегі етістіктерді зерттеуде назарға алынды.

Корпустық лингвистика туралы жалпы зерттеулер әлемдік тіл білімінде Е.Финеганның «Language: its structure and use» [321], Мак Энери және Э.Вильсонның «Corpus Linguistics» еңбегінде [322], В.Захаровтың «Корпусная лингвистика» оқу құралында [45], А.Н. Барановтың «Корпусная лингвистика», «Введение в прикладную лингвистику» [323], А.В.Зубов пен И.И.Зубованың «Информационные технологии в лингвистике» [324] еңбектерінде көрініс табады. Сондай-ақ әр тілдегі Ұлттық корпустар туралы жеке талдаулар З.А.Сиразитдиновтың «Моделирование грамматики башкирского языка» [325], В.З.Демьяниковтың «Англо-русские термины по прикладной лингвистике и автоматической переработке текста», Ч.Ф.Мейердің «English Corpus Linguistics» [326], М.Кито мен М.А.Риссаненнің «A Language in transition: the Helsinki Corpus of English texts» [327] зерттеулерінде жасалады.

Ал қазақ тіл біліміндегі компьютерлік лингвистиканың зерттелуі А.Жұбановтың «Основные принципы формализации содержания казахского текста» (2002 ж.) атты қазақ қолданбалы лингвистикасына қосылған елеулі үлес саналатын докторлық монографиясынан басталады. Сондай-ақ А.Жұбановтың «Қолданбалы тіл білімінің мәселелері», «Қазақ тексінің статистикасы» еңбектері, А.Жаңабекованың «Қазақ тілі мәтіндеріне морфологиялық белгіленім қоюдың ғылыми-тәжірибелік негіздері», «Аннотацияланған корпустардың ғылыми-практикалық маңызы», «Тілдік корпустардың ғылыми-зерттеу әлеуеті», Б.Карбозованың «Қазақ тілі мәтіндер корпусындағы омонимдер мәселесі», «Корпустық лингвистикадағы метабелгіленім ұғымы туралы түсінік», С.Құлмановтың «Қазақ тілі мәтіндері корпусының электрондық базасы жайында», «Корпусқа енгізілетін мәтіндердегі сөздерді морфологиялық талдау және оларға шартты белгілер қою принциптері», «Мәтіндер корпусы негізінде қазақ тілінің грамматикалық сөздігін түзудің практикалық аспектілері» сынды ғылыми мақалаларында аталған тақырып әр қырынан талданды.

Орыс тілінде «разметка» деп аталатын корпус термині қазақ тіл білімінде «белгіленім» деп аталады. («Белгіленім» термині – А.Жұбановтікі). Корпус – компьютерлік жадқа енгізілген мәтіндер жиынтығы ғана емес, сонымен бірге автоматты түрде түрлі лингвистикалық талдаулар жасайтын интерфейс пен жабдықталған құрылғы. Корпуста тілді компьютерге бағындыру мақсатында тілдің әр деңгейі бойынша автоматты түрде талдау жасау міндеті алға қойылады. Автоматты талдауға, әсіресе, агглютинативті тілдер ыңғайлы. Өйткені мұндай тілдерде түбір мен қосымшаның жігі анық көрініп тұрады. Ұлттық корпустардың барлығында осылай түбір мен қосымшаны автоматты талдайтын морфологиялық белгіленім программасы алғаш жасалған. ҚТҰК-да да морфологиялық талдауыш (анализатор) қазақ тіліндегі сөз формаларды автоматты түрде түбір мен қосымшаға бөліп, олардың сөз табына қатысын және қосымшалардың грамматикалық сипаттамасын шартты белгілер арқылы көрсете алады. Мысалы, «балаларға» сөзін морфологиялық анализатор төмендегіше талдайды:

Балаларға

Лемма бала

Морфология зт,лар/кж+ға/бс

Сурет 31. «Балаларға» сөзінің морфологиялық белгіленімі

Корпуста морфологиялық талдау ғана емес, тілдің басқа деңгейлеріне жасалатын тілдік талдаулардың барлығы қамтылуы тиіс. Мұндай жан-жақты лингвистикалық талдауды «терең аннотацияланған» корпус жасайды. Белгіленім түрлерін тіл деңгейлеріне сәйкес бөлуге болады: морфологиялық белгіленім, синтаксистік белгіленім, метамәтіндік белгіленім, сөзжасамдық белгіленім, лексика-семантикалық белгіленім, анафоралық белгіленім.

Метамәтіндік белгіленім – мәтін туралы берілетін ақпарат (мәтін авторы, атауы, жылы, стилі, жанры т.б.) экстралингвистикалық белгіленім деп аталады. Ал тіл деңгейлерінің лингвистикалық белгіленімдері интралингвистикалық

белгіленімге жатады. ҚТҰК-да метамәтіндік белгіленім 23 параметрмен жұмыс істейді.

Мұхтар Әуезов "Қараш-қараш оқиғасы"	
Автордың аты-жөні	Мұхтар Әуезов
Стиль	көркем әдебиет
Мәтін аты	Қараш-қараш оқиғасы
Таралым типі	кітап
Мерзімі (жазылған уақыты)	
Дереккөзі	Мұхтар Әуезов. «Әңгімелер жинағы» Алматы: 2014
Мәтін формасы	жазбаша
Мәтін тақырыбы	
Аудитория жасы	бейтарап
Субкорпус	көркем мәтіндер
Мәтін көзі	көшірме(сканер)
Мәтін хронотопы	
Автордың жынысы	ер
Автордың туған уақыты	1897
Сөзформа саны	21539
Жарық көрген уақыты	

Сурет 32. Метамәтіндік белгіленім үлгісі

Лексика-семантикалық белгіленім – сөздердің тақырыптық немесе лексика-семантикалық тобын көрсететін автоматты бағдарлама. Әлемдік корпусарда лексика-семантикалық белгіленім морфологиялық белгіленімге қарағанда кейінірек қолға алынған. Бұл аталған белгіленімнің күрделілігімен түсіндіріледі. Дегенмен Орыс тілінің Ұлттық корпусында семантикалық белгіленімді жасау үлкен қиындықтар тудырмаған. Себебі орыс тілінің сөздік құрамына негізделген 10 000 сөзден тұратын «Лексикограф» мәліметтер базасы 1992 жылдан жұмыс істеп тұр. Аталған жүйе Е.В. Падучеваның жетекшілігімен жасалған және ондағы етістіктер базасы аса бай. «Лексикографтың» артықшылығы – сөздерді семантикалық ерекшелігіне қарай талдауға мүмкіндік беруі. Жүйенің негізгі бағытын жоба авторлары былай көрсетеді: «Лексикограф» жүйесі лингвистикалық зерттеудің жоғары технологиялық құралы ретінде жасалған. Оның негізгі мақсаты – алдын ала берілген семантикалық параметрлер (ең алдымен таксономиялық) бойынша сөздердің тізімін алу және алынған лексикалық кластардың тілдік сипатын, нақтырақ айтқанда, тіркесімділігін, басқару модельдерінің түрлерін, лексикадағы жүйелі қатынасқа негізделген диалектикалық қозғалысын зерттеу» [328, 156 б.]. XX ғасырдың соңында жұмысын бастаған «Лексикограф» базасы корпус құрастырушыларға лексика-семантикалық белгіленім жасауда көп көмегін тигізді. Нәтижесінде Орыс тілінің Ұлттық корпусында лексика-семантикалық іздеу 2004 жылдың қазан айынан бастап іске қосылған: «Көлемінің шағындығына қарамастан, теориялық тұрғыдан «Лексикограф» мәліметтерді эксперименталды зерттеу

базасы лексика-семантикалық корпусты құрудағы рөлі зор және корпустың лексика-семантикалық сөздігінің толық негізін құрады деуге де болады» [328, 157 б.]. Орыс тілінің Ұлттық корпусында лексика-семантикалық белгіленім жасауда сөздерге негізгі үш белгісі бойынша сипаттама берілген. Олар: 1. *Деривациялық сипаты бойынша*: «диминутив», «аугментатив», «аттенуатив», «*nomen agentis*», «*nomen femininum*» т.б.; 2. *Лексика-грамматикалық категориялар бойынша*: сапалық, қатыстық (сын есім), деректі, дерексіз (зат есім) т.б.; 3. *Жеке семантикалық сипаты бойынша*: тақырыптық (таксономиялық) класс, «бағалау», каузативтілік (етістік және етістік негізді сөздер) т.б.

Етістік – есімдерден кейінгі екінші орында тұратын күрделі грамматикалық категория. Түркі тілдеріндегі есім-етістік омонимиясын зерттеуші түркологтар түркі тіліндегі кез келген түпкі түбір әрі затты, әрі қимылды атап, синкретті болып келетінін айтады. Сондықтан да белгілі түрколог А.М.Щербак етістік формаларының бастауында қимыл атаулары жатады дейді. Нақ осы қимыл атаулары есім-етістік синкреттілігі болған уақытында құбылыс пен процестерді тілдік құралдар арқылы атау барысында нақты сөйлеу үстінде сөйлеушінің іс-әрекетті орындаушыға және іс-әрекетке қарым-қатынасын білдірудің негізі қалана бастаған. Осылайша қимыл атаулары арқылы етістік формаларының семантикалық көптүрлілігі қалыптасқан [329, 96 б.]. Қазақ лингвистері етістіктерге ерекше назар аударып, көлемді зерттеулер жүргізген. Қазақ тіл білімінде етістік дәстүрлі принциптер тұрғысынан құрылымдық грамматикада және аспектуалдық тұрғысынан функционалды грамматикада қарастырылуда. Ал корпустық лингвистикада етістіктердің әрі грамматикалық, әрі лексика-семантикалық қыры ескеріліп, классификацияланады.

Лексика-семантикалық белгіленімнің ең маңызды бөлігі – сөз таптарын іштей семантикалық топтарға және субтоптарға жіктеу. Бұл процесс әр сөздің ішкі мағынасын тану негізінде жүзеге асады. Лексика-семантикалық белгіленім негізгі сөз таптарын қамтиды. Ғалым С.Исаев атап өткендей, «Тіл білімінің қай саласын алсақ та, оның зерттеу объектісі я тікелей, я жанама түрде сөзге келіп тіреледі. Өйткені сөзде фонетикалық қасиет (дыбыстық кешен) те, лексика-семантикалық қасиет (ішкі мағына) те, сөзжасам қасиеті (жаңа сөз я жаңа мағына) де, грамматикалық қасиет (морфологиялық құрамы мен құрылысы, сөздердің түрленуі, сол арқылы бір-бірімен байланысқа түсуі, алуан түрлі грамматикалық мағыналар білдіріп, сөйлем құрауы) те бір-бірімен қабысып, қабаттасып келіп отырады» [330, 5 б.]. Ғалым лексика-семантикалық жағын ішкі мағына деп атайды. Яғни лексика-семантикалық топтарға жіктеу тікелей бір сөзді екінші бір сөзден ажырататын дифференциалды семасын анықтай отырып топтастыруды білдіреді.

С.Исаев семантиканы лексикалық семантика және грамматикалық семантика деп бөліп қарастырады. Нақтырақ айтқанда, лексикалық мағынаға негізделсе, лексикалық семантика, ал грамматикалық мағынаға негізделсе, грамматикалық семантика болатындығына назар аударады. Корпуста – лексикалық семантика, яғни сөздің ішкі мағынасы назарға алынады. Лексикалық мағына (семантика) ...сөздің ұғымдық мағынасы, сөздік мағынасы, сондықтан жалқы болады [330, 287 б.].

Етістіктердің лексика-семантикалық мағыналарын ғалым М.Оразов төмендегідей ұстанымдар негізінде анықтайды: 1) бір лексика-семантикалық топтың құрамына енетін сөздер өзара бір мағыналық сыңар (семалар) арқылы байланысып жатуы керек; 2) бір лексика-семантикалық топтың компоненті болып саналатын сөздер өз дербестігін сақтауы керек; 3) лексика-семантикалық топтың құрамындағы сөздер бірі екіншісін мағыналық жағынан толықтырып, ұштасып жатуы керек; 4) антонимдік, синонимдік сыңарлар да белгілі бір лексика-семантикалық топтың аясында болады; 5) лексика-семантикалық топтар іштей семантикалық топтарға бөлініп, сатылай байланысып жатады; 6) лексика-семантикалық топтың құрамындағы сөздер бір сөз табынан болуы керек т.б.

Осы уақытқа дейін де бірқатар грамматистер етістіктерді мағыналарына қарай классификациялаған болатын. Мысалы, Ы.Е.Маманов етістіктерді сегіз топқа бөліп көрсетеді:

- 1) іс-әрекеттік қимылды білдіретін етістіктер (жазу, сызу, көру, жұлу т.б.);
- 2) қозғалыс-қимылды білдіретін етістіктер (жүгіру, еңбектеу, домалау);
- 3) қалып-күй процестерін білдіретін етістіктер (ауру, ұйықтау, отыру);
- 4) көңіл күйді білдіретін етістіктер (қайғыру, мұңаю, қуану);
- 5) сапалық белгінің өзгерісін білдіретін етістіктер (ұзару, қысқару);
- 6) еліктеу-бейнелеу етістіктері (жылмындау, қылмындау, пысылдау);
- 7) туу, өсу қимылын білдіретін етістіктер (балалау, құлындау, боталау);
- 8) субъективтік реңдік етістіктер (азсыну, кемсіну, көпсіну).

Берілген классификацияға сөйлеу, ойлау етістіктері енген. Сондай-ақ объектіге қатысты орындалатын әрекеттерден гөрі, субъектінің тікелей жасайтын әрекеттері назарға алынған. Сондықтан бұлай бөлу тілдегі барлық етістіктерді толық қамти алмайды. Ал ғалым А.Ысқақов «Қазіргі қазақ тілі» оқулығында «Етістіктің лексика-семантикалық сипаты» деген тақырыпта етістіктерді 11 мағыналық топқа бөледі:

- 1) амал-әрекет етістіктері (босат, көтер, күрес ки, сыз, өлше т.б.);
- 2) қимыл-қозғалыс етістіктері (ауна, аудар, домала, зыт т.б.);
- 3) қалып-сапа етістіктері (жат, жантай, тұр т.б.);
- 4) ойлау-сөйлеу етістіктері (айт, сөйле, де, ескер т.б.);
- 5) өсу-өну етістіктері (балала, жапырақта, гүлде т.б.);
- 6) бағыт-бағдар етістіктері (бар, кел, кет, әпер т.б.);
- 7) көңіл күйі етістіктері (жыла, қайғыр, өкін, күл т.б.);
- 8) бейнелеу-еліктеу етістіктері (жарқыра, күркіре, тарсылдат т.б.);
- 9) дыбыс-сес етістіктері;
- 10) көру-есту етістіктері;
- 11) мінез-құлық етістіктері.

М.Оразовтың жіктемесінде де етістіктердің 11 мағыналық тобы анықталады.

ҚТҰК мәтіндеріне лексика-семантикалық белгіленім жасау мақсатында 18 200 етістік жинақталған. Сондықтан мұндай көлемді тілдік бірліктерді тек сегіз немесе он бір семантикалық топқа бөлу жеткіліксіз әрі мағыналары толық ашылмауы мүмкін. Лексика-семантикалық белгіленімді жасаушы топтың жетекшісі проф. Б.Момынова бұл мәселеге тереңдеп, кіші топтардың санын

72-ге жеткізді. Ал орындаушы топ жұмыс барысында кейбір етістіктердің семантикасын дәл беру үшін лесика-семантикалық топтардың санын тағы да арттырды. Етістіктің ең негізгі мағынасын дәл беруге талпынатын мұндай кіші мағыналық топтарды субтоптар деп атауға болады. Бұл классификацияның басты артықшылығы әр етістіктің мәнін нақты әрі толық танытатын семантикалық тобы болуында. Кіші семантикалық топтарды Б.Момынова 15 ірі класқа (лексика-семантикалық топқа) біріктірді [331]. Олар:

- 1) *Адамның жай-күйі, қалып етістіктері;*
- 2) *Көңіл-күй етістіктер;*
- 3) *Ым-ишарат етістіктері;*
- 4) *Қимыл-әрекет етістіктері;*
- 5) *Күнделікті тұрмыста орындалатын әрекеттер;*
- 6) *Ауру, емдеу//емделуге байланысты етістіктер;*
- 7) *Ритуал, дәстүрге қатысты етістіктер;*
- 8) *Бір түстен екінші түске ауысу, басқа бір күйге ену;*
- 9) *Сөйлеу етістіктері;*
- 10) *Просодикалық етістіктер;*
- 11) *Қоғам мен мәдениетке қатысты іс-әрекеттер;*
- 12) *Ғылым-білім салаларына байланысты қимыл әрекеттері;*
- 13) *Әскери қимыл етістіктер;*
- 14) *Құрылыс-сәулетке байланысты қимыл-әрекеттер;*
- 15) *Табиғат: аң, құс, үй жануарларына қатысты әрекеттер.*

Кластарға біріктірудегі негізгі мақсат – топтарды визуалды түрде қолдануға, іздеуге және түсінуге ыңғайлы түрде беру. Мысалы, «Табиғатқа қатысты әрекеттер» лексика-семантикалық тобы төмендегідей субтоптардан және тілдік бірліктерден құралады:

1. *Өсімдікке, егіндікке байланысты орындалатын амал түрлері* (егу, шабу, өсіру, шөптеу, суландыру, майқандау, малалау, миюлату, отау, себу, септіру, т.б.);

2. *Үй жануарлары мен аңдардың жүрісі, қимылы* (шоқырақтау, жорғақтау, лоқылдау, көткеншектеу, саржелу, жорғалау (бауырымен), азулау, тістелеу, қандау, тырналау, қарқырату, дал-далын шығару, ұзу (сіңірін), шөгу, қорқырату, үйелеу, атылу, шиырлау, қорқаулану, лыпылдау, майпаңдау, мәлкілдеу, мөңку, мүйіздесу, сару, сауырлау т.б.).

3. *Мал бағу, күтуге байланысты қимыл-әрекеттер* (жүгендеу, арқандау, шідерлеу, жүгіндіру, білдіру (бас), шалмалау, күзеу, күлтелеу, кісендеу, төлдету, қожырау, қайыру, шайт-шайттау, құраулау, қоректендіру, жаю, қылшықтау, құдилау, міну, конжию, котандау, қылтанақтау, қозықұлақтану, майсыздандыру (сүтті), матау, тағалау, өмілдіріктету, өру, т.б.).

4. *Аңдар, үй жануарлары шығаратын дыбыстар* (қышқыру, кісінеу, мөңіреу, маңырау, маңырасу, бақылдау, шіңгірлеу, пысылдау, ырылдау, арқырау, шәңкілдеу, мәңкілдеу, мәуілдеу, осқыру, т.б.);

5. *Үй және жабайы құстар, жәндіктер шығаратын дыбыстар мен оларға қатысты орындалатын әрекеттер* (ақтару (қоясын), үйрету, шашу (жем), бастырту (ұя), ұшу, қону, сусылдау, ауыздандыру, ысылдау, жыбырлау, жиырылу, ызындау, саяқсу, сүрлеу, тамақтау, т.б.);

6. Саятшылыққа байланысты әрекеттер (салу (бүркіт), үркіту, қуалау, түсу (ізге), аңду, соғу (қамшымен) т.б.);

7. Табиғи апаттар мен құбылыстарға байланысты орындалатын қимыл-әрекеттер (малмандай болу, малшыну, мәнерлеу, партылдау, себезгілеу, сел жүру, тасқындау, сөну, т.б.);

8. Табиғат сыйын пайдалану (шайқау (май), консервілеу, масақтау, теру, жинау, мәйектену, самалдау, сәскелету, шомылу, т.б.);

9. Табиғаттағы аңдар мен құбылыстарға ұқсас қимылдар жасау (қорбию, тарбию, пашырлату, рауандау, сақырлау, сарқылдау, сидандау, сойдию, судырау, сумандау, сырию, т.б.)



Сурет 33. Семантикалық топтың негізі

ҚТҰК-да етістіктердің лексика-семантикалық топтары осылай көрініс таппақ. Ал жалпы семантикалық белгіленімнің әлемдік тәжірибесіне келсек, мысалы, Орыс тілінің ұлттық корпусында семантикалық белгіленім түрлері, ондағы лингвистикалық ақпарат әр сөз табының ерекшелігіне қарай беріледі. Орыс тілінің ұлттық корпусын жасаушылар етістіктерді төрт түрлі белгісі бойынша аннотациялайды:

1. Лексика-семантикалық топтар, яғни мағынасына қарай қозғалыс, физикалық әсер ету, жасау, жою, эмоция, сөйлеу сынды семантикалық топтарға жіктейді. Белгіленімде әр лексика-семантикалық топ ағылшынша атаулармен белгіленеді. Мысалы:

t:move – движение (бежать, дергаться, бросить, нести)

t:move:body – изменение положения тела, части тела (согнуть, нагнуться)

t:put – помещение объекта (положить, вложить, спрятать)

t:impact – физическое воздействие (бить, колоть, вытирать)

2. Каузативтілік. Төмендегідей каузативті немесе каузативті емес дегенді білдіретін белгі қойылады:

ca:caus – каузативные глаголы (показать, вертеть)

ca:noncaus – некаузативные глаголы (видеть, вертеться)

3. Көмекші етістіктерді сипаттау. Олар фазалық және каузативті көмекші етістіктерге жіктелген:

aux:phase – фазовые (начать, продолжать, прекратить)

aux:caus – служебные каузативные (вызвать, привести (к))

4. Сөзжасамдық белгілеуде етістіктер префиксті етістіктерге, семельфактив және екіншілік имперфективке ажыратылады:

d:pref – приставочные глаголы (забегать, оглядеть)

d:semelf – семельфактивы (кивнуть, чихнуть, боднуть, качнуться)

d:impf – вторичные имперфективы (-ива-, -ва-, -а-) (выпивать, вбивать).

Орыс тілінің Ұлттық корпусында 6 сөз табының лексика-семантикалық белгіленімі бар: зат есім, сын есім, сан есім, етістік, есімдік, үстеу. Зат есім жеке сөз табы түрінде емес, деректі, дерексіз, жалқы деп іштей үш түрге бөлініп беріледі. Тақырыптық топты «таксономия» деп атайды. Семантикалық фильтрлерді тандауда бірнеше лексика-семантикалық топтарды қатар белгілеу бір сөздің бойындағы түрлі мағыналық ренкті ескеруге мүмкіндік береді. Ал етістіктер таксономия, каузативтілік, көмекші етістіктер және сөзжасам деген төрт негізгі белгісі бойынша сұрыпталып, 18 ірі және 9 кіші лексика-семантикалық топқа жіктелген.

Кесте 28

Орыс тілінің Ұлттық корпусындағы етістіктің лексика-семантикалық топтары

№	Общие семантические группы	Частные семантические группы
1	Движение в том числе	Изменение положения тела, части тела
2	Помещение объекта	
3	Физическое воздействие	Создание физического объекта
		Уничтожение
4	Изменение состояния или признака	
5	Бытийная сфера	Существование
		Начало существования
		Прекращение существования
6	Местонахождение в том числе	Положение тела в пространстве
7	Контакт и опора	
8	Посессивная сфера	
9	Ментальная сфера	
10	Восприятие	
11	Психическая сфера в том числе	Эмоция
		Воля

12	Речь	
13	Поведение человека	
14	Физиологическая сфера	
15	Природное явление	
16	Звук	
17	Цвет	
18	Запах	

ҚТҰК-да етістіктерге лексика-семантикалық белгіленім жасауда орыс тілінің тәжірибесі ескерілді, ондағы әр сипаттаманың қазақ тілі етістіктерінің семантикасын ашуға қабілеттілігі назарға алынғанмен, каузативтілік және көмекші етістіктерге қатысты белгілеу қазақ тіліндегі етістіктерге сәйкес келмейтіні анықталды. Қазақ етістіктерінде каузативтілік семантикалық категориялардың да, лексика-грамматикалық категориялардың да қатарына жатпайды, яғни семантикаға емес, грамматикалық құрылымға негізделеді. «Мәжбүрле», «күште», «көндірі», «душар ет» секілді таза каузативті етістіктердің саны көп емес, керісінше, көбінесе каузативтілік кел-тір, ту-ғыз, жеткіз т.б. түрінде өзгелік етіс формалары арқылы көрінеді. Сондықтан каузативтілік «белгі-код» ретінде алынбайды. Префикс арқылы жасалатын етістіктер де қазақ тіліне тән болмағандықтан, көмекші етістіктерге қатысты сипаттама белгіленімге енбейді. Аталған белгілеулердің орнына етістіктер Ұлттық корпуста салт/сабақтылық және болымды/болымсыздық категориялары бойынша сипатталады. Бұл ҚТҰК-дағы лексика-семантикалық белгіленімнің негізгі ұстанымымен байланысты. Қазақ тіліндегі лексика-семантикалық белгіленімнің мақсаты – семантикаға басымдық беру және семантиканы басты критерий етіп алу арқылы етістіктердің мағыналық реңкін жан-жақты ашып көрсету, мәнін түсіндіру. Себебі семантика сөз мәнінің, синтактикасының, қолданысының өзегі болып табылады.

Орыс тілінің Ұлттық корпусының үлгісімен Ресей ғылым академиясында Қалмақ гуманитарлық зерттеу институты Қалмақ тілінің Ұлттық корпусын әзірлеген. Қазіргі таңда қалмақ тілінің екі корпусы бар: біріншісі – РМГУ аспиранты Э.Ванькаева құрастырған «Қалмақ корпусы», екіншісі – «Қалмақ тілінің Ұлттық корпусы». Қалмақ тілінің Ұлттық корпусындағы лексика-семантикалық белгіленім төрт түрлі бағыт бойынша метасипаттама береді: *1. Лексика-грамматикалық сипаттама, 2. Лексика-тақырыптық сипаттама, 3. Бағалауыштық сипаттама, 4. Деривациялық сипаттама.*

Орыс корпусынан Қалмақ корпусының айырмашылығы – тақырыптық топтарды барлық сөз таптарына ортақ беруі. Корпус құрастырушылары бұлай берудің мәнін төмендегіше түсіндіреді: «Біз де барлық сөз таптары үшін тақырыптық топтар ортақ деген пікірге қосыламыз. Біздің ойымызша, сөз семантикасы тілдегі константа болып табылады. Мысалы, жануарлар тақырыбына қатысты сөздер тобы кез келген тілде бар және әртүрлі сөз таптарындағы лексикалық бірліктерге ортақ сема бола алады. Оған қоса қалмақ тілінде изафеттік құрылымдар (модн гер «ағаш үй») кең тараған. Егер екі зат есім қатар келсе, онда оның алғашқысы

сөз табы жағынан сын есім болады. Модн «ағаш» сын есім де, зат есім де бола алады. Қалай келсе де «t:plant» (тақырыбы: өсімдік) деп белгіленеді». Барлық сөз таптарын ортақ тақырыптық топтарға жіктеу бәрін стандарт дайын үлгіге салу арқылы техникалық жұмысты жеңілдетеді, жеке зерттеуді қажет етпейді, бірақ «стандартты топтастыруда» сөз мағыналарын, мағыналық реңктерін ашу қиын. Сол үшін әр сөз табының лексика-семантикалық белгіленімі жеке қарастырылып, классификациялары бөлек жасалғаны жөн.

Деривациялық сипаттама қалмақ және орыс корпустарында шеттілдік терминдермен берілген: «Сөздің лексика-семантикалық сипаты сөзжасамдық белгіленіммен тығыз байланысты, бірақ басты мақсаты болып саналмайды. Сол себепті тек ең басты және жиі кездесетін деривациялық сипаттамалар берілді. Егер сөз «-го»-ға аяқталса, онда аталған лексикалық бірлік каритив деп белгіленді» – деген құрастырушылардың пікірінен деривациялық белгілеуді сөз семантикасы үшін аса маңызды деп санамайтындықтары байқалады. Осыны ескере келе, орыс және қалмақ корпустарындағы деривациялық сипаттаманы беру үлгісі Қазақ корпусында басқаша көрініс таппақ.

Өлемдік корпустардағы лексика-семантикалық белгіленімдерді талдай келе семантикалық және лексика-грамматикалық белгілеудің кең таралғанына көз жеткізуге болады. Бұл аталған сипаттамалардың лексика-семантикалық белгіленімнің өзегі болуымен байланысты. ҚТҰК-да да осы дәстүр сақталады. Сонымен қатар Қалмақ тілінің Ұлттық корпусындағы аталған сөздердің коннотациясына қатысты белгілеу түрі қазақ тіліне де аса қажет деп танылды. Себебі қазақ тіліндегі етістіктерде жеке тұрғанда көрінетін эмоциялық реңк бар, ол сөйлем және мәтін мазмұнына тікелей әсер етеді.

Шеттілдердің ұлттық корпустарындағы белгіленімді саралау нәтижесінде қазақ тіліндегі етістіктердің лексика-семантикасын жан-жақты ашатын белгілеу жүйесі әзірленді әрі қазақ тіліндегі белгіленім жасаудың негізгі шарттары анықталды. Қазақ тіліндегі етістіктердің лексика-семантикалық белгіленімі:

- тілдегі етістіктердің жалпы және жеке семантикасын анықтайды;
- лексика-грамматикалық сипаттама береді;
- коннотациялық реңкі анықталады;
- қазақ тілінің табиғатына сай семантикалық топтарға жіктеледі;
- кіші семантикалық топтар қолдануға ыңғайлы ірі кластарға біріктіріледі;
- корпустағы белгіленімдерді кодтауда көпшілікке түсінікті терминдер мен атаулар, ұғымдар қолданылады;
- етістіктердің омонимділігі ескеріліп, әр мағынасына жеке белгіленім жасалады.

Жоғарыдағы шарттарға сәйкес ҚТҰК-да лексика-семантикалық белгіленім төмендегідей 5 бағыт бойынша әзірленеді:

1. *Сөзжасамдық сипатына қарай:* дара/күрделі және негізгі/туынды. Сөз құрамы мен тұлғасын көрсету сөз мағынасын жақсырақ тануға мүмкіндік беретіндіктен, белгіленім құрамына қосылып отыр. Нақтырақ айтқанда, берілген етістіктің түбір етістік немесе басқа сөз негізінде туындаған туынды етістік немесе бірнеше сөзден құралған күрделі етістік екендігін білдіретін грамматикалық сипаты негізінде туынды мағына анықталады. Мысалы: «Ызында»

етістігі «ызың» еліктеуішінен жасалған туынды құрылым, сондықтан оның негізгі мағынасы – табиғаттағы жәндіктер шығаратын дыбысты білдіруі. Осы мағына негізінде оның лексика-семантикалық тобын айқындауға болады.

Ызыңда	етістік	дара	туынды ызың+да
--------	---------	------	-------------------

Сөзжасамдық белгі-код корпуста дәстүрлі терминдермен аталады, атап айтқанда негізгі, туынды және дара, күрделі болып түсінікті түрде жіктелінеді. Бұл мамандардың жаңа терминдерді меңгермеуінен немесе артқа тартуынан емес, керісінше, қолданушыға бұрыннан таныс ұғымдарды беру арқылы қолдануды барынша жеңілдету үшін жасалып отыр. Себебі «аугментатив», «аттенуатив» секілді ағылшын терминдерін лингвистер түсінгенімен, басқа саланың мамандары түсінбеуі мүмкін.

2. *Лексика-грамматикалық мағынасына қарай*: болымды/болымсыз, салт/сабақты. Бұлай беру аталған категориялардың әрі семантикалық, әрі грамматикалық сипатқа ие лексика-грамматикалық категориялар болуымен тығыз байланысты. Яғни, олар сөзге тек грамматикалық қана емес, семантикалық мән үстейді, мағынасына жанама әсер етеді. Ғалым М.Оразов та «Етістік» атты еңбегінде салттылық, сабақтылық категориясын семантикалық категория деп анықтайды [332, 158 б.].

3. *Жеке (кіші) семантикасына қарай* әр сөздің түпкі айырушы семасы анықталады, мағына терең танылады. Солардың негізінде 100 кіші семантикалық топтан тұратын классификация жасалды.

4. *Жалпы (ірі) семантикасына қарай* бөлу – кіші семантикалық топтарды ортақ семалары негізінде ірі кластарға біріктіру. Бұл іздеу процесін жеңілдетуге бағытталған. Орыс корпусының тәжірибесіне сәйкес, корпуста 5-6 ірі топтың берілуі қолданушының визуалды қабылдауына ыңғайлы. Қазақ тілі корпусында ондай ірі кластар саны – 15.

5. *Коннотативті сипатына байланысты* «жағымды», «жағымсыз», «бейтарап» деген үш белгі қойылады. Әр етістікте белгілі бір эмоциялық реңк болады және сол реңк оның мағынасына, контексте қолданылуына, синтактикасына тікелей әсер етеді. Кейбір синоним сөздер тек эмоциялық реңкі бойынша ғана ажыратылады. Ал кей етістіктер әр контексте әртүрлі реңк алуы мүмкін. Бұл негативті және позитивті мәнді ажыратуға ғана емес, тілімізде екі мәнде де қолданыла алатын етістіктерді анықтап көрсету арқылы қазақ тілінің мүмкіндіктерін таныту үшін қажет.

Етістіктердің лексика-семантикалық белгіленімін жасау процесі 7 кезеңнен тұрады. Олар:

- бірінші: қазақ тіліндегі дара етістіктерді жинастыру;
- екінші: басқа тілдер корпусындағы лексика-семантикалық белгіленім жүйесін, үлгілерін талдау, сараптау;
- үшінші: қазақ және шетел тіл біліміндегі етістіктің лексика-семантикалық классификацияларын жинақтап, саралау, соның негізінде нақты мағыналық топтарды анықтау;

- төртінші: жинақталған етістіктердің лексика-грамматикалық сипаттама-сын беру;
- бесінші: етістіктерді лексика-семантикалық топтарға жіктеу;
- алтыншы: жіктеуде қиындық тудырған етістіктер бойынша талдау жасау;
- жетінші: белгіленген алгоритм бойынша дайын мәліметтерді базаға енгізу, тексеру.

Лексика-семантикалық белгіленімнің әзірлемесін дайындауда практикалық жағына қатысты бірқатар мәселер туындады. Атап айтқанда, омоним және көпмағыналы етістіктерді контексте ажырату, бейнелеуіш мәндегі етістіктерді сипаттау, түрлі мағыналық реңкке ие етістіктерді нақты бір семантикалық топтың құрамына қосу және т.б. Аталған белгіленім ҚТҰК-да алғаш рет әзірленіп жатқандықтан, мұндай күрделі тұстарының болуы заңды. Алғашқы кезеңде мақалада көрсетілген базалық сипаттамаларды беру белсенді түрде жүзеге асырылады. Алдағы уақытта жоғарыда санамаланған мәселелердің оңтайлы шешімі ретінде семантика-синтаксистік модельдер жасауды қолға алуға болады. Мұндай модельдер әлем корпустарында да әлі толық жасалмаған.

Әлемдік корпус жасау тәжірибесі көрсеткендей, лексика-семантикалық белгіленім жасау – өте күрделі және мәтінді автоматты талдау үшін аса маңызды белгіленім. Етістіктердің лексика-семантикалық белгіленімін әзірлеу қолданушыларға мынандай мүмкіндіктерді ашады:

- Сөз мағынасын тереңірек тану;
- Кез келген уақытта электронды түрде қолдану;
- Синонимдес сөздерді анықтау;
- Тіл үйрену;
- Лексикографиялық еңбектер құрастыру;
- Семантикалық зерттеу жүргізу және т.б.

Лексика-семантикалық белгіленім болашақта мәтін талдау процесінде маңызды рөлге ие семантикалық-синтаксистік модельдер жасаудың алғашқы қадамы болмақ. Параллель электронды ресурста да семантикалық анализатордың болуы тілді әр қырынан қарастыруға, қазақ тілінің құрылысын, жүйесін жан-жақты зерттеуге жол ашады.

Зерттеушілер мен тілші мамандардың қажырлы еңбегінің арқасында күннен күнге дамып, қоры байып келе жатқан ҚТҰК тілдің мүмкіндігін танытатын, оны түсіну мен қолдануды жеңілдететін және зерттеуді жеделдететін қазақ тілінің басты цифрлы базасына айналуға.

11.4 Қазақ тілінің ұлттық корпусында етістіктің лексика-семантикалық белгіленімін қою жолдары

Қимыл-қозғалыстың, іс-әрекеттің, қалыптың, жай-күйдің, көңіл күйдің динамикалық үдерісін, түрлі құбылыстардың қозғалысын білдіретін сөз табын етістік деп атаймыз. Кез келген құбылыс, әрекет мезгілсіз, уақытсыз жүзеге аспайды. Уақыт, субъекті, объекті және қимылдан құралған құбылыстар күнделікті айналамызда болып жатады. Кез келген сөз табы секілді етістік те дара күйінде жалаң тұрып ешқандай грамматикалық та, лексикалық та мағына бер-

мейді. Белгілі бір сөзбен тіркесіп, сөйлем ішінде, құрмалас сөйлемдерді байланыстыруда түрлі грамматикалық мағынаға ие болады. Сөз таптарының ішіндегі грамматикалық мағына мен формаға бай сөз табының бірі болып саналады. Етістік басқа сөз таптарына қарағанда күрделісі, қолданыс аясы мен өрісі өте кең, себебі етістіктің лексика-семантикалық және лексика-грамматикалық сипатын анықтауда қатысатын субъектінің іс-әрекеті ғана емес, субъектіден тыс қоршаған ортадағы, қоғамдағы тілдік аядағы, ситуациялық жағдаяттарда орын алып жатқан сан түрлі қозғалыс, қимыл, әрекет, құбылыс, жай, күй сияқты үдерістер толық қатысады, сонымен қатар әр етістіктің семантикасын ашуда мезгіл, шақ ұғымдарын тыс қалдыруға болмайды. Себебі кез келген іс-әрекет белгілі бір уақыт шеңберінде жүзеге асып жатады. Ал лексика-грамматикалық сипатын анықтауда етістіктің грамматикалық формалары арқылы туындайтын грамматикалық мағыналарды айтамыз.

Жекелеген сөз таптарына лексика-семантикалық белгіленім енгізу тақырыбы қазіргі қазақ тілінің салалық дамуында, қазақ тілін автоматтандыруда, Қазақ тілінің ұлттық корпусын оның ішінде жекелеген ішкорпустар базасын құрудың негізінде өзектілігін жоғалтпайды.

Тіліміздегі метабелгіленім, белгіленім корпус, ішкорпус ұғымдары корпустық лингвистика саласының негізгі ұғымдары саналады. Жалпы корпус мәселесі әлемдік тіл білімінде неше жыл бұрын зерттеліп кеткен. Корпус ұғымына анықтама беріп статикалық талдау жүргізген шетелдік ғалымдар Ч.Ф.Мейердің, М.Кито мен М.А.Риссаненнің, Мак Энери және Э.Вильсонның еңбектерінен көруге болады [326], [327], [322]. Корпусты қолдану, құрастыру жайында мәліметті Э.Финеганның зерттеулерінде [321], корпус ұғымының тар мағынада жұмсалуды В.П. Захаровтың еңбегінде [45], корпустық лингвистиканың негізін мәтін құрайтындығы В.В. Рыковтың зерттеулерінде [333] сөз болады. Ал ҚТҰК-ның негізін салған А.Жұбановтың зерттеулерінен [158] компьютерлік лингвистика, метабелгіленім, белгі-код деген негізгі ұғымдарды көреміз. Ғалым қазақ сөздерінің картотекалық базасының негізінде автоматтандырылған сөз қорын жасауда, қазақ тілінің статистикалық мәліметтерін сөздіктер түрінде шығаруда, ҚТҰК-ның негізін салуда аянбай еңбек етті. А.Жұбановтың бастамасын ары қарай дамытып қазіргі ұлттық корпустың ішкорпустарын автоматтандыру ісіне белсене қатысып, корпуста енетін мәтіндерге метабелгіленім енгізу жұмыстарын ұйымдастырып, белгіленім қою мәселесіне зерттеу жүргізіп жүрген ғалым, профессор А.Жанабекованың еңбектерінде [12] осы мәселелер кеңінен сипатталады. Сондай-ақ жас ғалым докторант М.Иманғазина корпуста қатысты мақалаларында семантикалық белгіленім мен лексика-семантикалық белгіленімнің арасын ажыратып, күрделі тұстарын анықтап береді.

Қазақ тілінің грамматикасының тарихында етістіктің жұмсалуды өрісі, қолданысы, лексика-семантикалық топтары А.Байтұрсынұлынан бастау алып, кейіннен А.Аманжолов, Н.Сауранбаев, Ы.Маманов [334], М.Оразов [332], А.Ысқақов [310], Н.Оралбаева, А.Қалыбаева [335] т.б. ғалымдардың еңбектерінде егжей-тегжейлі жан-жақты қарастырылып зерттелді.

Қимыл-қозғалыстың қалпын, қарқындылығын, өту сипатын анықтайтын грамматикалық формалардың қатары көп. Сондықтан етістік сөз табы грам-

матикалық мағынаны анықтайтын грамматикалық формаларға өте бай сөз табы.

Жоғарыда аты аталған ғалымдардың зерттеулерінде етістіктің лексика-семантикалық топтарын мағынасына қарай бөліп берудегі негізгі ғылыми ұстаным түрліше сипат алды. Ал 1967 жылғы ұжымдық еңбек «Қазақ грамматикасында» [336] мағынасына қарай етістіктерді 6 ірі топқа бөледі және іштей кіші мағыналық бөліністерге бөледі. Ескертетін жайт, бұл ұжымдық еңбекте іс-әрекеттің, амалдың, жай-күй мен қалыптың қарқындылығын білдіретін етістіктер түбір күйінде еш қосымшасыз салт пен сабақтылық категориясына қарай ажыратылып берілген және көмекші етістіктерді де мағыналық топтың алтыншы түріне енгізген. Жалпы саны 6 топтан тұратын етістіктің лексикалық әрі мағыналық жігін ажыратып көрсеткен.

Қазақ тілінде етістіктер әртүрлі қалыпты, жағдайды, алуан түрлі әрекетті, қозғалысты білдіреді. Белгілі бір етістіктің мағынасын тек мәтін арқылы анықтай аламыз. Бір етістіктің өзі мәтіндегі мағыналық қолданысына қатысты әртүрлі ұғымда, тура мағынасымен қатар, ауыспалы мағынада жұмсалатынын білеміз. Мысалы кел етістігінің лексика-семантикалық тобын анықтайық: Жарқыным, айып емес кел дегенің, Өзің келсін деген соң, мен де келдім. *Келсем де досты танып, дұшпан көріп, Қасымнан алыста тұр*, келме дедім (Айттыс). Жастар, сендердің білімдерің көп. Оған дауым жоқ. Дегенмен біз алпысқа келдік. Сендер жиырмаға келдіңдер, соны ұмытпаңдар (Ж.Молдағалиев). Жас балалар үлкен адамдарға қарағанда әлдеқайда намысшыл, менменшіл келеді («Халық мұраты»). Үндемей кіріп келдің тайпақ қара, Өзің қандай, мен қандай байқап қара (Жамбыл, Шығармалар жинағы). Бірінші сөйлемде кел етістігі тура мағынасында белгілі бір орыннан бері қарай жүру, жақындау мағынасында яғни қимылдың субъектіге қатысын, қозғалысын білдіретін топқа жатады. Екінші сөйлемде кел етістігі ауыспалы мағынада, адам жасының белгілі бір мөлшерге жетуі, толуы сияқты ұғымды білдіреді. Келесі сөйлемде бол етістігінің мағынасында ауыспалы мағынада жұмсалса, соңғы сөйлемде негізі етістіктің жетегінде жұмсалып көмекшілік, қимылдың сипатын білдіріп тұр. Яғни тілде етістіктер өз мағынасынан бөлек ауыспалы мағынада да жұмсала береді, етістіктің лексика-семантикалық топтарын анықтау өте қиын, мүмкін емес, өйткені қимылды білдіретін сөз табы лексикалық мағынаға өте бай. Сондықтан ғалымдар ортақ мағынадан туатын сөздерді лексика-семантикалық сипатын анықтай отырып ірі топтарға, ірі топтарды ішінен микро топтарға ажыратып берген нұсқаларынан әртүрлілікті байқаймыз. Өйткені барлық етістіктерді мағынасына қарай ажырату өте қиынға соққан, дегенмен басты ерекшеліктерін анықтай отырып топтарға бөліп көрсеткен.

1962 жылы жарық көрген «Қазақ тілі» [337] мен 1967 жылы жарық көрген «Қазақ тілінің грамматикасы» еңбектерінде етістіктің лексика-семантикалық топтарға бөлінісі бірдей десек болады. Тек атауларында ғана өзгешелік бар. 1967 жылғы еңбекте етістіктің мағыналық бөлінісін салт пен сабақтылық қатынас негізінде анықтаса, 1962 жылғы еңбекте етістіктің жай-күйіне, қалпына, нысанына қатысты анықтағандығын және бұл еңбекте 7 топқа бөлгендігін байқаймыз [337]. Екі еңбекте де мағына жағынан бөлініс бірдей,

тек 1962 жылғы еңбектегі дерексіз мәнді етістіктер 1967 жылғы объектімен тікелей байланысты іс-әрекетті білдіретін сабақты етістіктермен пара-пар келеді. Екі еңбектегі етістіктердің қолданысы, семантикасы бірдей.

Ы.Маманов етістік категориясын негізгі етістік, функциялық етістік тұлғалар деп екіге бөледі. Негізгі етістіктің өзін өзара түбір етістік пен модификациялы етістіктер деп тағы екіге бөледі. Түбір етістік деп барлық етістіктің грамматикалық формаларына негіз болатын, лексикалық мағынаға ие сөздер, ал жасалу жолына қарай тағы да бірнеше түрге бөлінетіндігін айтады. Ғалым «Етістіктерді мағыналық ерекшеліктеріне қарай топтастырып бөлу өте қиын, тіпті мүмкін де емес. Дегенмен, оларды мағынасы жағынан басты ерекшеліктеріне қарай топтастыруға болады деп 8 топқа бөліп береді (29-кесте). Бірақ түбір күйінде етістіктерді мағынасына қарай топтастырып берудің грамматика үшін ешқандай практикалық мәні жоқтығын алға тартады. «Грамматика үшін түбір етістіктерді сыртқы тұлғалық бірлігіне қарай топтастырып бөлу қажет. Өйткені олар сыртқы тұлғалық бірлігі әртүрлі типтес сөздер тобын жасайды» – дейді [334]. Яғни түбір етістік өзі ғана тұрып лексика-семантикалық топ құрай алмайды, оған жалғанатын қосымшалардың негізінде («морфологиялық және синтаксистік тәсілдер») грамматикалық мағынаның үстіне лексикалық мағына жасау арқылы ғана мағыналық топқа бөлуге болатындығын айтады.

Кесте 29

Ы.Мамановтың жіктемесі

1	Іс-әрекеттік қимылды білдіретін етістіктер	жазу, сызу, қазу, жұлу, көтеру, тазалау, кесу, шашу
2	Қозғалыс қимылын білдіретін етістіктер	жүгіру, еңбектеу, домалау, жорғалау, өрмелеу, қозғалу, жылжу, қашу, жүзу
3	Қалып-күй процесін білдіретін етістіктер	ауыру, ұйықтау, тұру, отыру, жату, тынығу, жүдеу, жабығу, арықтау, семіру т.б.
4	Көңіл күйін білдіретін етістіктер	қайғыру, мұнаю, қуану, ренжу, күлу, жылау, ашулану, қызғану, т.б.
5	Сапалық белгінің өзгеру процесін білдіретін етістіктер	ұзару, қысқару, сарғаю, ағару, көбею, азаю, жұқару, қалыңдау, т.б.
6	Еліктеу, бейнелеу етістіктері	жымындау, қылмындау, пысылдау, ыңқылдау, гүрсілдеу, жылтылдау, бұртию, күржию, т.б.
7	Туу, өсу қимылын білдіретін етістіктер:	балалау, құлындау, қоздау, боталау, күшіктеу, жұмыртқалау, т.б.
8	Субъективтік реңді етістіктер. Бұлар сөйлеушінің қимылға субъективтік көзқарасын көрсетеді.	азсыну, көпсіну, кісімсу, ұлықсыну, білгішсіну, т.б.

М.Оразов «Қазақ тілінің семантикасы» атты еңбегінде сөздерді лексика-семантикалық жүйеге түсіру үшін олар парадигмалық қатынасқа негізделеді, осы топты құрайтын әрбір мүше (тілдік бірліктер) өзіндік мағыналық дер-

бестікке ие болып, сол жүйеде белгілі бір қызмет атқару қажет деп жазады [332]. Ғалым сема деген ұғымды қолданады. «Әр сөздің мағыналық құрамы іштей семалардан тұрады. Лексика-семантикалық топ құрау үшін, топ мүшелерінің лексикалық мағынасын анықтауда семаларға сүйенеміз» дейді. В.Г.Гак семаларды үшке бөледі: архисема, дифференциалды сема, потенциалды сема. Архисема сөздерді белгілі бір лексика-семантикалық топқа біріктірсе, дифференциалды сема сөздердің лексикалық мағынасына негіз болады деп көрсетеді. М.Оразов аты аталған семаларға қатысты қағидатты қолдай отырып, сөздерді лексика-семантикалық топқа топтастыруда сөздердің ортақ семасы жалпылауыш сема (архисема) дегенді, ал сөздерді лексика-семантикалық топтың басқа мүшелерінен ажыратып тұратын, сөздердің жеке дербестігін қамтамасыз етіп тұратын семаны айырушы (дифференциалды) сема деген атауды қолдайтындығын жазады.

Жалпылаушы семалар мен айырушы семалар сөздің лексикалық мағыналық тобын құрайды дедік, ал кейбір тілдік жағдаяттарда біреуі жетекші сема, екіншісі қосалқы көмекші сема қызметін атқарады дейді ғалым. Яғни М.Оразов сол кездің өзінде функционалды грамматика ұғымдарын, теориялық мәселелерін білген. Ядро, периферия сияқты ұғымдарды негізге алған. Сөз таптарын лексика-семантикалық топтарға топтастырғанда негізгі өзек (негізгі сема), периферия (қосалқы сема) деп қарағандығын байқаймыз. Ғалым етістік сөз табын лексика-семантикалық топтарға бөлу мәселесінде осы жалпылауыш семаны негізге алғандығын көреміз. «Жалпылаушы сема тематикалық, лексико-семантикалық топтарға сөздерді топтастыруға негіз болады, ал айырушы сема сол топтың мүшелерінің жеке-дербес жасауына жағдай жасайтын сема болып есептелінеді» - дейді [332]. М.Оразов аты аталған еңбегінде қазақ тіліндегі етістіктерді жалпылауыш семаларына қарай 12 топқа бөледі, атауыштық семасына қарай 45 топқа бөледі (30-кесте).

Кесте 30

М.Оразовтың топтастыруы

	Ірі мағыналық топтар	Ішкі мағыналық топтар
1	Амал-әрекет етістіктері	Қоректенуді білдіретін етістіктер: іш, же, ұртта, аса, дәм тат, сорапта, жалма, т. б.
		Жасау, өндіру етістіктері: жаса, соқ, жарат, тік, сал, құрастыр, қабы, т. б.
		Қирату, бүлдіру етістіктері: қират, бүлдір, бұз, сындыр, күл талқан ет, күлін көкке ұшыр т. б.
		Бөлу, ажырату етістіктері: бөл, ажырат, айыр, сұрыпта, т. б.
		Өлшеу етістіктері: метрле, қарыста, білемде, литрле, адымда, т. б.
		Белгілі бір құралмен орындалатын амал-әрекетті білдіретін етістіктер: балтала, бауызда, бұрғыла, пышақта т. б.

2	Қозғалу етістіктері	Жердегі қозғалысты білдіретін етістіктер: жүр, кел, еңбекте, сырғана, жорғала, т. б.
		Судағы қозғалысты білдіретін етістіктер: жүз, сүңгі, малты, жалда, т. б.
		Ауадағы қозғалысты білдіретін етістіктер: ұш, самға, қалықта т. б.
		Белгілі бір бағытқа қаратылған қозғалысты білдіретін етістіктер: кел, кет, бар, жақында, т. б.
		Бағытсыз қозғалысты білдіретін етістіктер: арала, шарла, кез, қыдыр т. б.
		Объектіден алыстауды білдіретін етістіктер: кет, алыста, ұза, без, т. б.
3	Қарым-қатынас етістіктері	Теңестіру мәнді етістіктер: қос, теңестір, салыстыр, бала, т. б.
		Меншіктеу, тәуелдеу етістіктері: иемде, меншікте, жіп тақ, ие бол, т. б.
		Амандасу, қоштасу етістіктері: амандасу, сәлемдесу, көрісу, танысу, қоштасу, қайырласу, т. б.
		Құрметтеу, қадірлеу етістіктері: құрметте, қадірле, ізетте, құрмет көрсету, т. б.
		Бастыру, бағындыру етістіктері: басқар, әмір жүргізу, сөз өткізу, т. б.
		Бағыну етістіктері: көну, бас иіу, ырқына көну (журу), бағыну, т. б.
4	Сезіну етістіктері	Жалпы сезіну мәнді етістіктер: сөз, байқа, абайла, аңғар, пайымда, т. б.
		Көру етістіктері: көр, көз сал, үніл, т. б.
		Дәм сезу етістіктері: кермектену, ащылтымдану, т. б.
		Иіс сезу етістіктері: бықсы, мұңкі, аңқы, т. б.
		Терімен байланысты сезіну етістіктері: сыздау, ашу, қышу, тызылдау, т. б.
		Есіту етістіктері: тыңда, есіт, құлақ сал, т. б.
5	Ойлау етістіктері	жалпы ойлау мәнді етістіктер (ойла, байымда, ой жүгірт, т. б.), жорамалдау, болжау етістіктері (болжа, бағдарла, тұспалда, межеле, т. б.)
		қорытындылау етістіктері (қорытындыла, шешімге кел, бел байла, тәуекел ет, т. б.)
		сенім етістігі (сен, дос деп сана, дос көр, жақын сана, т. б.)
		еске сақтау етістіктері (еске ал, зердеңде сақта, жадына алу, т. б.)
6	Сөйлеу етістіктері	хабарлау етістіктері (айт, жеткіз, хабарла, т. б.).
		сөйлеу етістігі (сөйлесу, әңгімелесу, әзілдесу, т. б.).
		сұрау-жауап етістіктері (сұра, жауап бер, жауап қайтар, т. б.)
		эмоционалды сөйлеу етістіктері (мыңқылда, көкі, отта, т. б.),
7	Дыбыс-сес етістіктері	тарсылда, сатырла, тырс етті, гүрсілде.

8	Эмоционалды етістіктер	Таңдану етістіктері: таңдану, таңырқау, аузы ашылу, т. б.
		Қуану етістіктері: шаттану, қуану, мәз болу, т. б.
		Қанағаттану етістіктері: рахаттану, қанағаттану, көңлі толу, т. б.
		Қайғы-мұң етістіктері: капалану, қайғыру, қабырғасы қайысу, т. б.
9	Қалып-сапа етістіктері	Созылыңқы амалды білдіретін етістіктер: отыр, тұр, жатыр, жүр, т. б.
		Зағтардың түсінің өзгеруімен байланысты қолданылатын етістіктер: қарай, сарғай, қызар, т. б.
		Зағтардың көлемінің өзгеруімен байланысты қолданылатын етістіктер: шөк, тапалдану, мықыраю, т. б.
		Зағтардың сандық мөлшерінің өзгеруімен байланысты қолданылатын етістіктер: азай, көбей, молай т. б.
		Адамның не зағтардың ішкі қасиеттерінің өзгеруімен байланысты қолданылатын етістіктер: ірі, шірі, үгіл, тоз, т. б.
		Әдеттенуге байланысты етістіктер: әдеттену, үйрену, дағдылану, т. б.
10	Өсіп-өну етістіктері	ту, төлде, қозыла, ботала, т. б.
11	Субъектив реңктерді білдіретін етістіктер	кісімсі, адамсы, аспанда, баласын т. б.
12	Табиғи құбылыстарға байланысты етістіктер	азына, гуілде, борандату, найзағай түсу, т. б.

Етістіктің лексика-семантикалық топтастыру жайында зерттеу жүргізіп, ғылыми негіздеп берген профессор А.Ысқақовтың тұжырымдамасын ерекше атауға болады. Ғалым түбір етістіктердің семантикалық жүйесін мағына жақындығына сай 11 топқа бөледі (31-кесте). Қимыл атауын білдіретін етістікті лексика-грамматикалық белгілері жағынан бірыңғай жеке сөз табы ретінде саналатындықтан, олардың өздерін іштей бір-біріне мағына жағынан жақындықтарына, өзара қызметі жағынан байланысына орай топ-топқа бөлуге болатындығын жазды.

А.Ысқақов та профессор Ы.Мамановтың сөзін қолдағандай. Себебі етістік те басқа сөз таптары сияқты мағыналық бөлініске бөлуге болады. Бірақ грамматика үшін етістікті лексикалық мағынасына қарай бөлу «сондайлықты ұтымды бола қоймайды» деген еді. Себебі етістік жалаң күйінде, дара, түбір күйінде тұрып, ешбір синтетикалық, аналитикалық тәсілдерсіз оның грамматикалық формасы мен мазмұны осы тәсілдер арқылы анықталмай, семантикалық құрылымына орай жіктеп алмай, лексика-семантикалық топқа бөлуге болмайтындығын айтады. Өйткені грамматикалық мағына грамматикалық формалар арқылы жасалатындықтан әрі сөздің құрылысы мен құрамына байланысты болғандықтан, лексикалық семантикасын анықтау мүмкін емес. Ы.Маманов секілді А.Ысқақов та түбір күйіндегі етістіктерді мағына мен өзара қызметі жағынан орайластырып топтастырады [310].

А.Ысқақовтың жіктемесі

1	амал-әрекет етістіктері	(босат, көтер, күрес, ки, сыз, өлше т. б.
2	қимыл-қозғалыс етістіктері	ауна, аудар, домала, жыт, қаш, секір т. б.
3	қалып-сапа етістіктері	жат, жантай, тұр, тос, күт, ұлғай т. б.
4	ойлау-сөйлеу етістіктері	айт, сөйле, де, ескер, жатта т. б.
5	өсу-өну етістіктері	балала, жапырақта, гүлде, өс, өн т. б.
6	бағыт-бағдар етістіктері	бар, кел, кет, найт, әпер, әкет т. б.
7	көңіл күйі етістіктері	жыла, қайғыр, өкін, күл, нуан т. б.
8	бейнелеу-еліктеу етістіктері	жарқыра, күркіре, дүркіре, тарсылда, сыртылда, сырқыра т. б.),
9	дыбыс-сес етістіктері	
10	көру-есту етістіктері	
11	мінез-құлық етістіктері	

Зерттеуші А.Хасенова да етістіктерді әр жақты қарап, түбір етістік дейтін межеге сыйғызуға келетін сөздерді іс-әрекет, амал, қалып процесінің динамикасын білдіруі жағынан төмендегіше 6 лексика-семантикалық топқа жіктеп көрсеткен [335]. Олар: объектімен тікелей байланысты іс-әрекетті білдіретін етістіктер; субъектінің қозғалысының беталысын, бағытын білдіретін етістіктер; субъектінің қалпының сапасының өзгеру процесімен байланысты айтылатын етістіктер; бейнелеу етістіктері; ішкі объектілі салт етістіктер; көмекші етістіктер. Жоғарыда аты аталған ғалымдардың бөлінісі бір-бірімен мазмұндас. Лексика-семантикалық белгіленімге етістіктің мағыналық бөлінісін осы ғалымдардың классификациясын салыстыра отырып іріктеп, макро және микро топтарға жіктеу арқылы сұрыпталған нұсқасын ұсынамыз.

А.Байтұрсынұлы атындағы Тіл білімі институты 2022-2023 жылдар аралығында «Мемлекеттік тілдің ақпараттық-инновациялық базасы ретіндегі Қазақ тілінің ұлттық корпусын әзірлеу: ғылыми-зерттеу және оқыту интернет-ресурсы» атты бағдарламалық-қаржыландыру жобасы аясында ҚТҰК-ның іш-корпустарын жасап жатыр.

Грамматика бөлімі 18200 етістіктің лексика-семантикалық белгіленімін енгізу жұмыстарын қолға алды. Лексика-семантикалық белгіленім бағдарламасын жасауға қажетті қазақ тіліндегі етістіктердің лексика-семантикалық топтарының тізімін жинақтап берген профессор Б.Момынованың 72 түрлі классификациясын негізге ала отырып етістіктерді морфологиялық, сөзжасамдық, лексика-грамматикалық, лексика-семантикалық белгі-кодтарға жіктеп, автоматты бағдарламасын әзірлеу үстінде. Мысалы, «ызалану, ашуға булығыу» мағынасында жұмсалатын доңайбаттан етістігінің лексика-семантикалық белгіленімін ішкорпусқа енгізіп, лингвистикалық автоматтандыру жұмысының тілтанымдық ресурсын жасап көрсетсек. Жалпы етістіктің лексика-семантикалық белгіленімі 6 параметр бойынша анықталады.

1. Морфологиялық белгіленім: сөз табы – етістік.
2. Сөзжасамдық белгіленімнің өзі іштей төрт белгі-код арқылы анықталады:
 - а) сөздердің морфемалық құрамы: доңайбат+тан;
 - ә) сөзжасамдық тәсіл: синтетикалық;
 - б) тұлғасына қарай – дара;
 - в) құрамына қарай – туынды.
3. Лексика-грамматикалық белгіленімнің өзі іштей екі белгі-код арқылы жасалады:
 - а) іс-әрекеттің болымды/болымсыздығы – болымды;
 - ә) салттылық/сабақтылық қатынасқа қатысты – салт.
4. Лексика-семантикалық белгіленім бойынша екі белгі-кодпен өлшенеді:
 - а) макротоп – **көңіл күй етістіктері**;
 - ә) микротоп – адамның сезім-күйін білдіретін етістіктер тобын құрайды.
5. Коннотациясы – жағымсыз.
6. Иллюстрациясы – «Қанша елемейін деп сырттай доңайбаттанғанмен, іші құрмағыр тағат таба алмайды» («Қазақ әдебиеті»). Осындай үлгіде әр етістіктің лексика-семантикалық, лексика-грамматикалық белгіленімін енгізіп шығамыз. Етістік күрделі сөз табы екендігін ескере отырып, көпмағыналы қимыл атауларының метабелгіленімін қою ерекшелігін қарастырсақ. Мысалы, үш түрлі мағынада жұмсалатын егіл етістігінің лексика-семантикалық белгі-кодтарын кесте түрінде берсек. 1. Бір нәрсенің отырғызылып, өсірілуі мағынасындағы егіл етістігінің жұмсалуды. 2. Белгілі бір аурудың таяқшасын қанға жіберіп індеттен сақтандыру деген мағынадағы медициналық термин ретінде кездесетін егіл етістігінің ерекшелін атауға болады. 3. Іші-бауыры елжіреу, көңілі босау, езілу ұғымында жұмсалатын егіл етістігінің үшінші коданысын атап кетуге болады.

Кесте 32

Көп мағыналы етістіктердің ішкорпуста берілу үлгісі

Белгіленім түрлері	Ішкі белгі-кодтар	ЕГІЛ 1	ЕГІЛ 2	ЕГІЛ 3
Морфологиялық белгіленім	Сөз табы	етістік	етістік	етістік
	Сөздің морфемалық құрамы	Ек+іл	Ек+іл	Егіл
Сөзжасамдық белгіленім	Сөзжасамдық тәсіл	синтетикалық	синтетикалық	синтетикалық
Лексика-грамматикалық белгіленім	Тұлғасы	туынды	туынды	негізгі
	Құрамы	дара	дара	дара
	Іс-әрекеттің болымды/болымсыздығы	болымды	болымды	болымды
	Салттылық/сабақтылық	салт	салт	салт

Лексика-семантикалық белгіленім	Макротоп	күнделікті тұрмыста орындалатын әрекеттер	ауру, емдеу// емделуге байланысты етістер	адамның жай-күйі, қалып етістіктері
	Микротоп	Құрал-сайман, жабдықтар арқылы орындалатын әрекеттер мен мөлшерді білдіретін етістіктер	Ем-дом қабылдау	Адамның бір күйден екінші күйге ауысуы, біреудің екінші біреуді әлдебір күйге түсіруі, күй-қалыптың ауысуы
	Коннотациясы	бейтарап	бейтарап	бейтарап
	Иллюстрация	Ауа райы бүгін шынында да рақат, жайлы еді. Емхана ауласындағы асфальт жолдар бойына егілген кара ағаштар қанатын жайып, айналасына салқын саясын түсіріп тұр (О.Сәрсенбаев, Жиде).	Мұның үстіне бруцеллезге қарсы шығарылған вакцина адамға мезгіл-мезгіл егіліп тұрса, нұр үстіне нұр (Ә.Дайырбеков, Малдан адамға жұғатын аурулар).	Елжіреп, егіліп сөйлегенде анаға пар келетін, оған жететін пенде жаралмаған (С.Хайдаров, Жүрек сыры).

Бірнеше мағына білдіретін етістіктер қатарын 1, 2, 3 деген рет санын бере отырып, белгіленімін сол рет бойынша жеке-жеке енгіземіз (ЕГІЛ 1, ЕГІЛ 2, ЕГІЛ 3). Ең бірінші өзінің тура өз мағынасында жұмсалыуына мән береміз, келесі екінші, үшінші ұғымдарын мағына жақындығына сәйкес ажыратамыз.

Байқағанымыздай, етістіктерді лексика-семантикалық классификация мен белгі-кодтар арқылы одан ары тереңдей анықтау үшін әр етістіктің макро, микро топтары яғни семантикасы анықтала отырып, олардың лексикалық мағынасын, омонимдік, ауыспалы, синонимдік қатарын ажырату мақсатында иллюстрациялық материал бір сөйлеммен берілді. Морфологиялық белгіленімде етістіктің құрамын (түбір мен қосымшасын) анықтау, тұлғасына қарай дара/күрделі, болымды/болымсыз, негізгі/туынды, субъектінің қатысына қарай салт/сабақты, коннотациясына қарай жағымды/жағымсыз, бейтарап екендігін анықтау әрі сол негізде белгіленім қою бұрын-соңды атқарылмаған жұмыс. Тек біраз етістіктерге морфологиялық талдау жасалып, ұлттық корпусқа белгіленімі енген болатын. Бұл жаңа белгіленім түрі болып саналады. Етістіктер тек мағыналық жағынан ажыратылып, бір-біріне жақын ұғымдағы, бірдей қызметтегі қимыл атаулары ірі және кіші мағыналық топтарға жіктеліп қана қоймайды, сонымен қатар әр етістіктің лексика-грамматикалық белгіленімі, морфемдік белгіленімі, құрамы мен тұлғасына қарай бөлінісі, етістің қатыстылығына қарай яғни субъектінің ырықсыз, ортақтық, өзгелік семасының берілуіне байланысты субъектінің тікелей өзі не субъектісіз атқарылуына қарай белгіленім қойылады. Сөзжасамдық белгіленім де қойылады. Әріп реті-

мен берілген етістіктердің синтетикалық, аналитикалық тәсіл арқылы жасалуы да ескеріледі.

Сонымен, лексика-семантикалық ішкорпус базасына енгізуге 18 200 етістік жинақталып, мағыналық реңктері анықталды. Gmail платформасында етістіктердің лексика-семантикалық белгіленімін әзірлеу үшін алты түрлі белгі-код енгізу жұмыстарынан ақпарат берілді. Бірінші: морфологиялық белгіленім бойынша сөздің құрамы мен сөз табы анықталды, екінші: сөзжасамдық сипатына қарай дара, күрделі; негізгі, туынды, үшінші: лексика-грамматикалық белгіленім негізінде салт, сабақты; болымды, болымсыз; төртінші: коннотативтік сипатына қарай жағымды, жағымсыз, бейтарап деп жіктелді, бесінші: етістіктердің лексика-семантикалық бөлінісі іштей ірі және кіші топтарға бөлінді, алтыншы етістіктің мағынасын дұрыс анықтау мақсатында мысалы, яғни иллюстрациясы қоса берілді.

Грамматика бөлімінде лексика-грамматикалық, лексика-семантикалық, сөзжасамдық сипаттағы етістіктің белгі-кодтарын қою арқылы жалпы ұлттық корпусстың ішкорпусында автоматты бағдарламасы дайындалып жатыр. Бұл тілді цифрландыру, автоматтандыру мақсатында үлкен жетістік саналады. Лексика-семантикалық белгіленім енгізу жалпы тілші қауымға, ЖОО ұстаздары мен мектеп мұғалімдері, филологтарға грамматикалық білімдер жүйесін жан-жақты зерттеуге мүмкіндік туғызары анық.

11.5 Корпуста күрделі етістіктерді автоматты тану мәселелері

Табиғи тілді өндеуге арналған арнайы компьютерлік бағдарламалардың ішінен мәтінге белгіленім енгізуді автоматты түрде жүзеге асыратын бағдарламалар айырықша орын алады. Корпустарға белгіленім жүргізу (tagging, annotation) ең бір күрделі жұмыс деуге болады, әсіресе қазіргі кездегі корпустар көлемінің үлкендігі бұл жұмысты одан сайын күрделендіре түседі. Морфологиялық талдауды (тэггерлер) автоматтандыратын компьютерлік бағдарламалар жұмысының нәтижесінде мәтін бойындағы әрбір лексикалық бірлікке олардың грамматикалық сипаттамалары «тіркеліп» беріледі. Демек, **белгіленім** жүргізу дегенді мәтін бірліктеріне алдын ала шартты түрде арнайы қабылданған тэгтерді «тіркеп» жазу әрекеті деп ұғыну қажет.

Морфологиялық белгілер тек сөз таптарын ғана анықтап қоймайды, ол сонымен бірге сол сөз табына ғана тән грамматикалық категорияларды да қамтиды. Морфологиялық анализатор сөздерге морфологиялық талдау жасауда оларды жеке-жеке сөздерге бөліп тастайды, содан кейін барып әр сөзге тиісті сөз табы мен грамматикалық тэгтерін қояды. Күрделі етістіктерге қатысты мәселе осыдан туып отыр.

Күрделі етістіктер құрамы екі етістіктен тұрады, бірақ олардың сыңарлары сөйлемде жеке дара мағына білдірмейді. Сондықтан күрделі етістіктерді біртұтас тілдік бірлік ретінде танитын автоматты программа қажет болады.

Корпуста сөздердің лексикалық мағыналары көрсетіліп тұрады. Ал күрделі етістіктерге келгенде жеке сыңарларға ажырап кететіндіктен, олар білдіретін біртұтас мағынаны беру мүмкіндігі болмайды. Бұған қоса қазақ тілінде

күрделі етістіктердің мағыналары берілген арнайы түсіндірме сөздігі жоқ, олар түсіндірме сөздіктерде сөздік ұяларына енбеген.

Қазіргі қазақ тілінде күрделі етістіктерге қатысты әлі күнге шешімін таппаған мәселелер бар. Күрделі етістіктерге жақын тілдік құбылыс – аналитикалық етістіктер екені белгілі. Аналитикалық етістіктер тілде сөзжасамдық бірлікке жатпаса да, лексикалық және грамматикалық мағынаның қосындысы ретінде біртұтас қимылды білдіреді. Сондықтан олардың да сыңарларын бөліп жаруға келмейді. Корпуста морфологиялық анализатор программасына да күрделі етістіктер сияқты аналитикалық етістіктерді де танитын мүмкіндік туғызу керек және түсіндірмелерін беру қажет. Біз жұмысымызда күрделі және аналитикалық етістіктерге қоса құранды етістік және етістікті тұрақты тіркестердің де мағына тұтастығына, яғни сыңарлары қосылып, бір ғана қимылды білдіретінін негізге алып, бұларды **күрделі құрамды етістікті сөз тіркестері** аясында қарастырамыз.

Күрделі құрамды етістікті сөз тіркестерін лексикографиялау, яғни сөздіктерінің болуы мен оларды корпуста біртұтас тілдік бірлік ретінде танудың шешімін табуы өзге ұлт өкілдерінің қазақ тілін үйренуінде, жалпы тіл меңгертуде маңызды. Күрделі құрамды етістікті сөз тіркестері өзге тілдерге аударылғанда түрлі аударма жұмыстарын жүзеге асыруда өте қажет.

Құрамы мен мағына тұтастығы жағынан **күрделі құрамды етістікті сөз тіркестеріне** мыналар жатады:

- 1 Аналитикалық формалы етістіктер;
- 2 Күрделі етістіктер (сөзжасамдық);
- 3 Құранды етістіктер;
- 4 Етістікті тұрақты тіркестер;

Аналитикалық формалы етістіктер құрамы екі не үш етістіктің тіркесуі арқылы жасалады. Мысалы, *жасай сал, жүгіре жөнел, айта баста* т.б. Бірінші сыңарлары толық мағыналы негізгі етістік, екінші сыңарлары – тіркес құрамында лексикалық мағыналарынан айрылып, грамматикалық мағына білдіруге көшкен көмекші етістіктер. Мысалы: *жасай сал, жүгіре жөнел, айта баста* т.б.

Күрделі етістіктердің (сөзжасамдық) құрамы жағынан аналитикалық формалы етістіктерден айырмашылығы тек екі етістіктің (үшеуі тіркеспейді) тіркесуі арқылы жасалады. Сөзжасамдық тұрғыдан танылып жүрген күрделі етістіктерде (*кіріп шықты, әкеліп берді*) екі сыңары да лексикалық мағыналарын жоғалтпайды, екі етістіктің мағынасы қосылып, қимылдың екі кесіндісінен тұратын біртұтас қимылды білдіреді. Ғалымдар осы ерекшелігіне қарай күрделі етістіктерді сөзжасамдық құбылысқа, сөзжасам саласының нысанына жатқызады. Мұндай күрделі етістіктер де сөйлемде бір ғана сұраққа жауап беріп, бір сөйлем мүшесінің қызметін атқарады.

Құранды етістіктер аналитикалық және күрделі (сөзжасамдық) етістіктерден айырмашылығы – бірінші сыңары есім сөзден жасалады. Күрделі етістіктер сияқты құранды етістіктердің екінші сыңары – етістіктер лексикалық мағынасын жоғалтпайды, есім сөзбен тіркесіп, біртұтас қимыл мағынасын білдіреді, яғни есім сөзге тіркесіп, оны етістікке айналдырады (дедек қақты, еңбек етті, еңбек сіңірді, үлгі қылды, қор болды, мазақ қылды, ән салды т.б.).

Етістікті тұрақты тіркестер лексикалық қабаты жағынан фразеологизмдерге жатады. Етістікті тұрақты тіркестер екі немесе үш, кейде төрт компонентті болып келе береді: *бұлжытпай орындады, дәнігіп алды, жазбай таныды, төбе шашы тік тұрды, өкшесі жерге тимеді* т.б. Мұндай түйдекті тұрақты тіркестер компоненттер ортасына сөз салуға болмайды. Олар тілдің даяр қолданылатын тілдік бірліктері болып табылады.

Сонымен, **күрделі құрамды етістікті сөз тіркестері** дегеніміз – ең кемі екі сыңардан жасалған, біртұтас мағына білдіретін, сөйлемде бір ғана сұраққа жауап беріп, бір ғана сөйлем мүшесінің қызметін атқаратын идиомаланған және идиомаланбаған тіркестер.

Біздің күрделі құрамды етістікті сөз тіркестеріне жатқызып отырған түрлері жеке дара алғанда өзіндік мағыналық ерекшелігі бар әртүрлі тілдік деңгей элементтері болғанмен, қимылды білдіретін біртұтас тілдік құбылыс болуымен бір-бірімен сабақтасады. Біздің оларды сабақтас тілдік құбылыс ретінде қарастыруымыздың өзіндік себептері бар. Біріншіден, тілді өзге ұлт өкілдеріне оқытуда осы аталған күрделі құрамды етістіктерді түйдекті тіркестер ретінде қарастырып, мағыналарын түсіндіру қажет; Екіншіден, аударма жасауда да мұндай күрделі құрамды етістіктердің арнайы сөздігі болмағандықтан, олардың мағыналарын жеке сыңарларының мағынасы бойынша сөзбе-сөз аударуға тура келеді. Әсіресе, келешек машиналық аударма жасауда мұндай күрделі құрамды етістіктердің аударма сөздігін пайдалану қажеттілік тудырады. Қазіргі интернет жүйесіндегі қазақша-орысша не орысша-қазақша аудармалардың қате болуы осындай алғышарттардың жасалмауынан болып отыр. Үшіншіден, күрделі құрамды сөз тіркестерінің ішіндегі сөзжасамдық күрделі етістіктер мен аналитикалық формалы етістіктер түсіндірме сөздіктерге енбеген. Олардың тізімделген жеке сөздігі де жоқ. Төртіншіден, мәтінді автоматты тануда сөз құрамына морфологиялық талдау жасайтын анализаторлар да мұндай сөздік жоқ болғандықтан, олардың құрамындағы сөздерді жеке-жеке талдап көрсетеді де, олардың бір сөйлем мүшесі ретіндегі біртұтас мағынасы ашылмай қалады.

Қазақ тілінде ойды жеткізудің негізгі құралы етістіктер болғандықтан, олардың бірнеше етістіктен құралып бір ғана қимылды білдіруі жиі кездеседі. Сондықтан, қазақ тілін үйренушіге, әсіресе, етістік сөз табының күрделі құрамды бірліктерінің мағынасын да білу маңызды болып табылады.

Н.Оралбаева, Б.Қасым пікірлері бойынша, қазақ тіліндегі қимыл мәнін білдіретін күрделі сөздерді етістіктің аналитикалық форманттары мен күрделі етістіктерге бөліп тану керек [314]. Бұл пікір қазақ тіл білімінде қолдау да тауып келеді. Ғалымдар пікірінше, оқып шықты – етістіктің аналитикалық форманты болса, кіріп шықты – күрделі етістік. Алғашқысында оқу қимылының өту сипатын білдірсе, екіншісінде кіру және шығу қимылынан тұратын біртұтас әрекетті білдіреді дейді. Етістіктің аналитикалық форманттары грамматикалық мағына үстеуіне қарай қазір функционалды грамматикада қимылдың өту сипаты (аспектуалдылық) категориясының (өрісінің) тілдік құралдары болып табылады. Демек, қимыл мәнді күрделі сөздердің ішінде лексикалық және грамматикалық мағына үстеуіне қарай әртүрлі құбылыстар бар екендігін

көреміз. Осы орайда күрделі етістіктердің сөзжасам мен формажасамға қатысы түбегейлі шешімін таппай келеді.

Қазақ тілі сөздіктерінде күрделі етістіктер сөздіктердің ешбіріне енбеген. Сондықтан біз жоғарыда берілген етістікті тіркестердің лексикографиялануын басшылыққа алып, көмекші етістіктер/тірек компоненттер арқылы жасалған күрделі етістіктерді лексикографиялау мүмкіндіктерін қарастырамыз. Төменде бірнеше күрделі етістіктің лексикографиялық құрылымына тоқталмақпыз. Ал жинаған материалдарымыздың толық көлемін өзіміз құрастыратын «Сөздікте» (Қосымшада) береміз. Біз күрделі/аналитикалық етістіктерді тірек компоненттер бойынша «Қазақ тілінің ұлттық корпусынан» жинадық. Қазіргі кезде жиналған күрделі етістіктер/аналитикалық етістіктер саны – 2000-нан аса.

Енді етістікті тұрақты тіркестердің лексикографиялануындағыдай, күрделі етістіктерді лексикографиялауда да соған ұқсас принциптер сақталады.

Күрделі етістік жасауда да екінші сыңары көбінесе тірек компоненттер деп атаған көмекші етістіктер болып келеді. Алайда күрделі етістіктерде тірек компоненттердің мағынасы өздерінің лексикалық мағынасы шеңберінде қолданылады. Абстракты әмбебап мағынаға айналмайды. Мысалы:

Алып шықты. к ү р д. а) Затты, нәрсені қолына алып не бірге ертіп сыртқа шығарды. Ал енді ыдыс-аяқ салынған сандық ашылғанда бір қаробка пышақтарды а л ы п ш ы қ т ы (Қ.Өзенбаева, Асықты көрпе). Мәди қызды байлаудан босатып есік алдына а л ы п ш ы қ т ы да, шашынан сипап көзінің жасын сүрте бастады (Ә.Әбішев, Жаралы сұңқар). ә) а у ы с. Бір жұмысты өзі не өздерінің күшімен бітірді, нәтижеге қол жеткізді. Жарыста өз командасын қалайда а л ы п ш ы ғ а т ы н (Ауызекі).

Барып шықты. к ү р д. Бір жерде болып, қайта кетті, біреудің үйіне соғып кетті.

Кіріп шықты. к ү р д. Бір жерге жолай соғып кетті. Үлкен кісі үйіне к і р і п ш ы қ т ы (Қ.Түменбаев, Ағаш ескерткіш).

Айтып келді. к ү р д. Бір нәрсе туралы хабардар етті, естігенін келген жеріне жеткізді.

Барып келді. к ү р д. Бір жерде болып, қайта оралды, жолсапардан қайтты. Жаппар мынау қолға ұстар дәнеңесі жоқ мөп-мөлдір, шегіне шетіне жетіп болмайтын тұлдыр көкке қарай б а р ы п к е л д і (Ә. Кекілбаев, Мұнара).

Әкеліп берді. к ү р д. Бір затты алып, біреуге ұстатты.

Апарып берді. к ү р д. Бір затты біреуге не бір жерге жеткізді. Онда біздің жұмысымыз жоқ, жоғалыңдар, - деп апайлар дорбаны апарып берді (Қ.Өзденбаева, Додаға түскен дорбалар).

Жоғарыда *шық, кел, бер* көмекші етістіктері арқылы жасалған күрделі етістіктер берілген. Аталған көмекші етістіктер немесе басқа да көмекші етістіктер/тірек компоненттер күрделі етістіктер (сөзжасамдық) жасағанда, біртұтас қимылдың кесіндісі ретінде көбінесе өздерінің тура мағынасында жұмсалады.

Қазақ тілін өзге ұлт өкілдеріне оқытуда аналитикалық формалы етістіктердің де түсіндірме, аударма сөздіктерін құрастыру керек болады. Өйткені аналитикалық форманттар, көмекші етістіктер мағынасы грамматикаларда,

оқулықтар мен оқу құралдарында жазылып, оқытуда үйретіліп жатқанмен, қазақ тілін оқыту әдістемесінде көмекші етістіктер семантикасы түсінуге, түсіндіруге қиын күрделі мәселе болып отыр. Осы орайда аналитикалық формалы етістіктер тілде сөзжасамдық құбылыс ретінде танылмаса да, оның лексика-грамматикалық мағынасын түсіндіретін құрал (сөздік) дайындау тіл үйренуде, сонымен қатар әртүрлі аударма жұмыстарын жүзеге асыруда маңызы зор. Мысалы,

Танысып шықты. а н а л. Бір нәрсе туралы ақпарат алуды аяқтады, бітірді. Тұрғанбике: Ий, Тоқтағали. Сенің жеке іс-қағаздарыңмен т а н ы с ы п ш ы қ т ы м (М.Ақмырзаұлы, Партизан).

Шолып шықты. а н а л. Бір қарап өтті. Жамбастап жатып ап, өзі екі ай тұрған мекенді қасқыр көзімен бір ш о л ы п ш ы қ т ы (Қ.Түменбаев, Қара қазан).

Оқып шықты. а н а л. Оқып бітірді, аяқтады. Текежан ол берген бір жапырақ қағазды ішінен о қ ы п ш ы қ т ы да, тісін шықыр еткізді, бірақ өңінде өзгеріс жоқ (Ә.Әбішев, Жаралы сұңқар).

Шық етістігі негізгі етістіктерге тіркесіп, үш аналитикалық формалы етістікте де іс-әрекеттің біткенін, аяқталғанын білдіретін мағына үстейді.

Бастала қойды. а н а л. Көп күттірмей, тез арада басталып кетті, тез-ақ басталды. Билеттегі орынды апайға беріп, «маған құдай жар болсын» деп қасына жалп ете қалдым, көпжасағыр концерт де көп ұзамай б а с т а л а қ о й д ы (Қ.Өзденбаева, Алып-ұшқан көңіл-ай!).

Сала қойды. а н а л. Оп-оңай орналастырды, жеңіл, тез-ақ салды. Бұрын 20 теңге беріп, арбамен әкеліп, автобуска с а л а қ о я т ы н заттарын енді таксимен тасып, торыққан жұрттың жан айқайын кімге айтарын білмей, өздерімен-өздері шер тарқатып, басқа елдің артынан қарап мен тұрмын (Қ.Өзденбаева, Арман мен Алдан).

Әкеле қойды. а н а л. Жеп-жеңіл әкеле салды, тез-ақ алып келді. Екі кесесін балаларыма ала салайын деп, сумкадан ақша алам дегенше апай лып етіп барып, қалтаға салып ә к е л е қ о й д ы (Қ.Өзденбекова, Коридордағы кемпір).

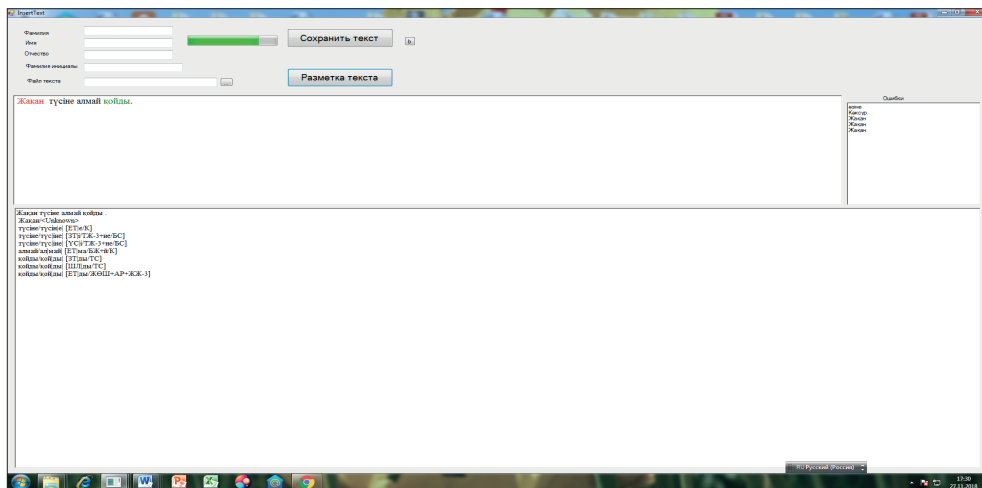
Қой көмекші етістігі жоғарыда берілген үш аналитикалық формалы тіркесте негізгі етістікке іс-әрекеттің тез арада әрі жеңіл-желпі орындалу мағынасын үстейді.

Ағыла бастады. а н а л. Үздіксіз келе бастады, көп келетін, баратын болды. Оның бұл бұйрығы да бұлжыған жоқ, жұрт ол хабарды ести сала, лег-легімен а ғ ы л а б а с т а д ы (Ә.Әбішев, Жас түлектер).

Ажырата бастады. а н а л. Екі немесе бірнеше затты бөле бастады, жігін аша бастады. Енді отыра қалып, едендегі алақандай жылтыр төсеніштерді пышақ ұшына оп-оңай іле көтеріп, бір-бірінен а ж ы р а т а б а с т а д ы (Қ.Түменбаев, Ағаш ескерткіш).

Сонымен, жоғарыда берілген материалдардан күрделі етістіктер мен аналитикалық формалы етістіктерді де осылайша мағыналарын ашып, сөздіктер құрастыруға болатындығын көреміз. Жоғарыдағы берілген мысалдар сияқты көптеген тіркестерді (күрделі етістік және аналитикалық формалы етістік) мағынасын түсіндіріп, мағыналық ерекшеліктерін айқындай беруге болады.

Корпуста морфологиялық анализатор күрделі етістіктерді жеке дара бөліп талдау жасайды. Мысалы: *келіп қалды, оқып шықты, жүгіре жөнелді* сияқты күрделі етістіктер кездескенде бұлардың әрқайсысын дербес етістік түрінде бөліп алады. Яғни, автомат компьютерлік бағдарлама мұндай етістікті тұрақты етістіктің (көзін сатты, дедек қақты т.б) аналитикалық формалы етістіктің, құранды етістіктің, күрделі етістіктің біртұтас тілдік бірлік екенін ажырата алмайды. Осымен байланысты құрылымы жағынан осындай күрделі етістіктерді программаға тізімдеп беру қажет.



Сурет 34. Корпустағы морфологиялық анализатордан өткен күрделі етістік

Жақан [ЗТ]

түсіне/түсін/е [ET/e/K]

алмай/ал/май [ET] ма/БЖ+й/К

қойды/қой/ды [ET] ды/ЖӨШ+AP+ЖЖ-3

Көріп отырғанымыздай, *түсіне алмай қойды* деген күрделі құрамды етістікті тіркесті (аналитикалық формалы етістікті) морфологиялық анализатор әрбір сөзін жеке-жеке бөліп алып, құрамына морфологиялық талдау жасаған.

Күрделі етістіктерді автоматты корпустық белгіленімге енгізу үшін алдымен күрделі құрамды етістіктерді ең алдымен тізімін жасап, түсіндірме сөздігін құрастыру қажет. Жоғарыда күрделі құрамды етістікті тіркестерді лексикографиялаудың үлгілері берілді. Келешекте мұндай сөздік корпустың морфологиялық анализатор жұмысын жетілдіруге, күрделі құрамды бірліктерді/ күрделі етістіктерді біртұтас сөз ретінде тануға мүмкіндік береді.

11.6 ҚТҰК-дағы сөзжасамдық белгіленім әзірмесі

Қос сөздерді кодтау және автоматты түрде белгілеу жолдары

Қазақ тілінде түбір сөздер негізгі түбір және туынды түбір сөздер болып бөлінеді. Мұның қай түрі де лексикалық бірлік ретінде сөздіктерге реестр

ретінде енеді, мысалы, негізгі мал сөзі де, туынды малшы сөзі де дербес лексикалық бірлік, сондықтан екеуі де реестрлік тізімге кіреді. Мал сөзі сияқты қазақ тіліндегі негізгі түбір сөздерді сөзжасамдық анализатор реестр тізімінен іздейді. Малшылық деген туынды сөз кездескенде оны реестрдегі тізімге қарап отырып, оның малшы сөзінен туып отырғандығын, ал малшы сөзінің реестрде тағы кездесіп тұрған мал сөзінен туып отырғандығын көрсетеді. Яғни сөздің соңынан бастап, реестр тізімінде бар сөздермен салыстыра отырып, сөздің түпкі түбіріне дейін бөлшектеуі қажет. Демек, сөзжасамдық анализатор үшін де сол тілдегі түбір (реестр) сөздердің тізімін лингвистикалық база ретінде беру қажет.

Қазақ тілінде сөздер жоғарыдағы малшы сөзіндегідей, тек қосымша жалғану арқылы ғана жасалмайды. Сондықтан сөзжасамдық белгіленім бағдарламасында, яғни сөзжасамдық анализаторға сөзжасамның басқа да тәсілдері туралы лингвистикалық нұсқаулық беріледі.

Қазақ тілінде туынды сөздер үш түрлі тәсіл арқылы жасалды: 1) *лексика-семантикалық тәсіл*; 2) *синтетикалық тәсіл*; 3) *аналитикалық тәсіл*.

Автоматты талдаулар үшін лексика-семантикалық тәсіл ескерілмейді, өйткені бұл тәсілмен жасалған сөздердің формальдық белгілері жоқ, сөз мағына кеңеюі арқылы жаңа мағыналарға ие болады. Корпус үшін сөзжасамдық белгіленім бағдарламасында синтетикалық және аналитикалық тәсіл арқылы жасалған туынды сөздер негізге алынады. Мұның ішінде синтетикалық тәсіл бойынша жасалған сөздерді талдауда программа реестр тізімі мен сөзтудырушы қосымшалар тізіміне сүйенсе, аналитикалық тәсілде көбінесе біріккен сөздер мен қос сөздердің тізіміне негізделеді, яғни программаға осындай сөздердің тізімі лингвистикалық база ретінде беріледі.

Аналитикалық тәсіл арқылы жасалған сөздерді күрделі сөздер дейміз. Мұндай күрделі сөздер: қос сөздер, біріккен сөздер, тіркесті түбір сөздер, қысқарған сөздер. Қысқарған сөздерді танудың формальді белгісі бас әріптермен сөздің бастапқы әріптерінің жазылуы немесе нүктемен сөздің толық берілмеуі. Программа бұл сөздерді формальді белгілері арқылы да табуына болады.

Қос сөздерді автоматты танудың бірнеше жолы бар. Біріншісі – қос сөздердің сыңарлары арасына қойылатын дефис белгісі. Мәтінде іздестірілген немесе мүлде басқа сөз іздестіруде конкарданса шыққан сөйлемдердегі қос сөздерге курсорды апарғанда, лингвистикалық белгіленімдер терезесінде ол сөздің қос сөз екендігі көрсетіліп тұрады, мұндайда программадағы іздеу дефис белгісі арқылы қос екендігін таниды. Сол үшін автоматты тану бағдарламасына «дефис белгісімен келетін сөздерді қос сөз деп тану қажет» деген нұсқау берілуі қажет. Қазақ тілінде дефис белгісі кейде сызықшаның орнына қойылып кететін жағдайлары мәтіндерде көп кездеседі. Әрине бұл – дұрыс емес. Программа мұндай жағдайда шатасып, кез келген дефис қойылған жерлердегі екі қатар сөзді қос сөз ретінде тануы әбден мүмкін. Сондықтан, корпуста салынатын мәтіндер техникалық түзетуден мұқият өткізілуі керек.

Сонымен қатар қос сөздерді автоматты тануда оларды тек қос сөздер ретінде көрсету аз, өйткені оның екі түрі бар: қайталама және қосарлама қос сөз болып бөлінеді. Бұл екі түрі де дефис арқылы жасалғанмен, қайталама қос сөздер

түбірдің қайталанып келуі арқылы жасалады, ал қосарлама қос сөздерде екінші сыңары басқа сөз болып келеді.

Егер қос сөз құрамында екінші сыңарда бірінші сыңардағы сөзді қайталанып қолданса, программа оны қайталама қос сөз ретінде тануы қажет және соған сәйкес «қайталама қос сөз» деп көрсетеді. Мысалы: қайта-қайта, әлсін-әлсін т.б. Белгіленім моделі: туынды, күрделі, анал., қос сөз, қайталама. Мұндағы туынды – жасалу тәсіліне қарай туынды сөзге жататындығы, күрделі – құрамына қарай күрделі сөз болып табылатындығы, анал. деген – аналитикалық тәсілмен жасалғандығы, қос сөз – сөздің күрделі сөздің қос сөз түріне жататындығы, қайталама – қос сөздің ішінде қайталама қос сөз болып табылатындығы.

Мұндай белгіленім қайталама қос сөздердің барлығына ортақ болып келеді. Ал егер қос сөздің екінші сыңары бірінші сыңарынан мүлдем басқа сөз болса, яғни алдыңғы сыңарды қайталамаса, программа оны «қосарлама қос сөз» деп тануы қажет. Мысалы: әке-шеше, ата-ана, бала-шаға т.б. Белгіленім моделі: туынды, күрделі, анал., қос сөз, қосарлама.

Демек, алдыңғы мен соңғының айырмашылығы – соңғы қайталама мен қосарлама деп ерекшеленуінде ғана.

Біз жоғарыда қос сөздерді табудың бірінші жолын айтып отырмыз. Екінші жолы – қайталама қос сөздер мен қосарлама қос сөздердің тізімін программаға лингвистикалық база ретінде беру. Мұндайда программа алдыңғы жолындай сөздердің екінші сыңарының қайталану/қайталанбай сипатына назар аудармайды, бұл жолы қос сөздің екі түрінің тізіміне ғана сүйеніп, мәтінде кездескен қос сөздерді ажыратады. Мұндай бағдарламада базаға берілген қос сөздердің тізіміне енбей қалған қос сөз белгісіз болып қалады, сөзжасамдық белгіленім қойылмайды. Сондықтан қос сөздерді автоматты тануда осы аталған екі жолды да программаға енгізу қажет, яғни программа қос сөздерді осы екі жолмен де қатар іздеп көру керек.

Қос сөздерді автоматты тану үшін қазақ тіліндегі қос сөздер тізімін академик Ә.Қайдардың қос сөздер сөздігін пайдаландық [338]. Қос сөздерді қосарлама және қайталама қос сөздерге бөліп, техникалық түзетуден өткіздік және қос сөздің екі тобына жоғарыда берілген белгіленімдерді кодтап енгіздік. Түзетуден өткізілген, кодталған тізімді программиске дайындап бердік. Берілген қос сөз саны – 4200.

Біріккен сөздерді кодтау және автоматты түрде белгілеу жолдары.

Қос сөздерді формальді белгілері бойынша анықтағандай, компьютерлік бағдарлама біріккен сөздерді де тануына болар еді. Біріккен сөздер екі сөздің бірігуі арқылы жасалады.

Автоматты танудың бірінші жолы – реестр сөздер тізіміне сүйене отырып, егер біріккен сөз құрамында екі дербес сөз қолданылып тұрса, мұндай сөздерді біріккен сөздер ретінде көрсетеді. Мәселен, саңырауқұлақ, итмұрын, аққала т.б. Модель: туынды, күрделі, анал., біріккен сөз.

Мұндағы туынды – жасалу тәсіліне қарай туынды сөзге жататындығы, күрделі – құрамына қарай күрделі сөз болып табылатындығы, анал. деген – аналитикалық тәсілмен жасалғандығы, біріккен сөз – сөздің күрделі сөздің

біріккен сөз түріне жататындығы. Ал кіріккен сөздерді бұл жолмен тану қиын, өйткені кіріккен сөздердің құрамы кірігіп, екі сөзге аражігі бөлінбейді, шегарасы көрінбейді, сондықтан кіріккен сөздерді программаға тек тізімдеп қана беру мүмкіндігі бар. Мысалы: бүгін, бүрсігүні, энеугүні т.б. Қазақ тілінде мұндай кіріккен сөздер саны көп емес. *Модель*: туынды, күрделі, анал., кіріккен сөз.

Біріккен сөздерді автоматты танудың екінші жолы – қос сөздердегідей олардың түрлеріне қатысты тізім жасау.

Біріккен сөздерді кіріккен сөздерден басқа, өз ішінде жасалу тәсіліне қарап:

а) екі сыңары да қазақ сөзі болатын біріккен сөз – қазақ тілінің өз ішіндегі сөздердің бір-бірімен бірігіп, бір ұғымды білдіруі арқылы жасалған сөздер. Мысалы, аққала, ұзынкұлақ, көкқұтан, *ақбөкен, ақжусан, атқосшы, итмұрын, саңырауқұлақ* т.б.

ә) төл будан – біріккен сөздің бір сыңарының басқа тілден енген сөз болып келуі арқылы жасалған сөздер. Мысалы: автотұрақ, машинажасау, телеайтыс, телеорталық, полимәдени, морфомүсін, миротізбек, метафизикалық, лингвомәдениеттану, инфрақұрылым т.б.

б) кірме будан – біріккен сөздің екі сыңарының басқа тілден енген сөз болып келуі арқылы жасалған сөздер. Мысалы: авиакасса, совхоз, партком, *геоакустика, гигабайт, кардиохирургия, лингвостатистика, макроэволюция* т.б. Бұлар көбінесе термин сөздер болып келеді.

Біріккен сөздердің бұдан басқа сөзалды, сөзсоңы біріккен сөздеріне жіктеп көрсете беруге болар еді. Біз әзірге жоғарыда аталғандармен шектеп отырмыз. *Модель*: туынды, күрделі, анал, төл будан/кірме будан.

Осылай біріккен сөздерді автоматты тану үшін қазақ тіліндегі біріккен сөздерді «Қазақ әдеби тілінің сөздігінің» 15 томының әрбір томының реестрінен бөліп алдық. Берілген біріккен сөздер саны – 6600.

Қосымша арқылы жасалған туынды сөздерді кодтау және автоматты түрде белгілеу жолдары.

Тілдегі синтетикалық тәсіл арқылы жасалған сөздерді дара туынды сөздер дейміз. Синтетикалық тәсілде негізгі түбір сөздерге сөзжасамдық жұрнақтар жалғанып жаңа туынды сөз жасалады.

Бұл тәсілмен жасалған туынды сөздерді сөзжасамдық анализатор реестр тізіміндегі сөздерді және сөзжасамдық қосымшалардың тізімі мен сипаттамасы берілген кестені пайдаланып таниды. Біріккен сөздерді анықтауда программа сөз құрамындағы бөліктерді реестрден іздеп тапса, қосымша жалғану арқылы жасалған сөздерде де осы іздеу тәсілін қолдануға болады, яғни сөз құрамында бір ғана толық мағыналы сөз тұрса, толық мағыналы сөзден кейінгі жалғанып тұрған реестрде жоқ бөлік қосымша болып шығады. Мысалы: малшы, малшылық, балалық, адамгершілік, абыройсыздық, балала, оюла, бірінші т.б. сөздерде мал, малшы, бала, адам, абырой, бала, ою, бір сөздері реестрде бар толық мағыналы сөздер. Ал осы сөздерге жалғанып тұрған *-шы, -лық, -гершілік, -сыздық, -ла, -інші* бөліктері дербес сөздер емес, сондықтан программа оларды біріккен сөздер ретінде танымайды, екінші кезекте сөзжасамдық қосымшалар кестесінен осы қосымшаларды іздестіреді. Программа мұндай іздеуде әр сөз табының сөзжасамдық қосымшалар тізімі берілген кестеге сүйенеді.

Есімнен зат есім тудырушы жұрнақтар

Реті	Қосымша	Мысал	Сипаттама	Шартты белгі Обозначения
1	-шы	қойшы, сақшы	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
2	-ші	тілші, әдебиетші	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
3	-шылық	малшылық, балықшылық	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
4	-шілік	егіншілік, тіршілік	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
5	-гершілік	адамгершілік, жаугершілік	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
6	-лық	жаңалық, байлық	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
7	-лік	өткірлік, қателік	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
8	-дық	батылдық, жылдамдық	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
9	-дік	мүмкіндік, шешендік	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
10	-тық	бастық, қастық	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж
11	-тік	биіктік, төстік	Есімнен зат есім тудырушы жұрнақ	Ес.зт.туд.ж

Қосымшалар кестесінен табылмаса, оларды тағы да қосымша бөліктеріне бөлшектеп көреді, мысалы: *-сыздық* қосымшасы кестеде жоқ, сол себепті анализатор оны *-сыз және -дық* деп тағы бөлшектеп, тағы да қосымшалар тізімінен іздестіреді. Әрі қарай іздестіру осылайша жалғаса береді. *Модель*: туынды, дара, синт. жұрнақ

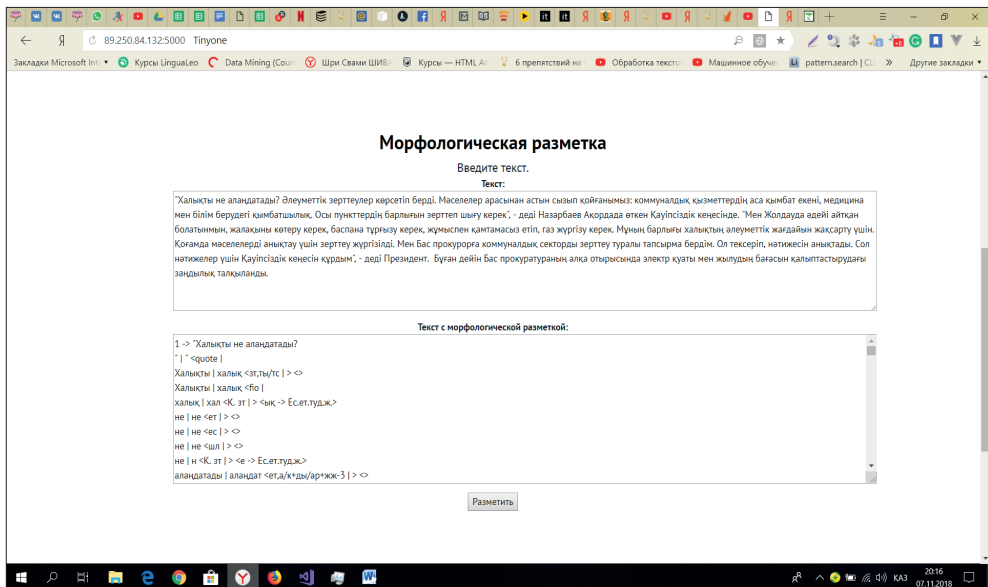
Сөзжасамдық жұрнақтарды әр сөз табына қатысты қысқаша шартты белгілермен беруге де болады. Мысалы: етістіктен етістік тудыратын жұрнақтар – Ет.ет.туд.; есімдерден зат есім тудырушы жұрнақтар – Ес.зт.туд. т.б.

Программада морфологиялық анализатор мәтіндегі сөздің негізін және оның қай сөз табына жататындығы реестр тізімі арқылы табады. Содан кейін табылған сөзден басқа екінші бөлігі қосымша болып анықталады. Қосымшаларды сөзжасамдық жұрнақтар кестесінен іздестіреді. Қосымша сөзтүрлендіруші қосымшалар кестесімен де салыстырады. Сөйтіп, сөзжасамдық қосымшалар кестесінен анықтап алғаннан кейін, оларға кестедегі шартты белгіні (кодты) қояды.

Сөзжасамдық белгінім қоюда туынды сөздермен қатар негізгі сөздерді түпкі түбір екендігін көрсетуге болады. Негізгі түбірлер реестр тізімінде

тұрған ешқандай қосымша жалғанбаған немесе ешқандай түбір сөз қосылып жазылмаған сөздің түпкі бөлігі. Мұндай негізгі түбірді – Т деп таңбалауға болады. Сөзжасамдық анализатор алдымен сөздердің осындай түпкі бөлігін іздеп тауып алуы қажет. Туынды сөз болмаса, ол негізгі түбірді бірден – Т (түбір) деп белгілеп, ал егер туынды сөз болса, оның негізгі түпкі түбірін бөліп алып, оған Т белгісін қойып, әрі қарай туынды сөзді жасаушы сөзжасамдық жұрнаққа қарай ығысады.

Ал төменде осы сайттағы белгілеу көрсетілген. Мұнда берілген мәтін сөйлемдерге және жеке сөздерге бөлінген. Ол сөздерге морфологиялық және сөзжасамдық белгіленімдер қойылған.



Сурет 35. Морфологиялық және сөзжасамдық белгіленімдер сайтының жұмысы

Қорыта келгенде, корпус мәтіндеріндегі сөздерге автоматты түрде сөзжасамдық белгіленімдер қоятын программа жасау ұлттық корпус құрастыруда үлкен бір нәтижелі көрсеткіштердің біріне жатады. Бұл лингвистикалық зерттеулерде сөзжасам мәселелерін зерттеуде таптырмас ақпараттық-анықтамалық құрал болады. Бұл сөзжасамдық белгіленім қою программасын 100 пайыздық шынайылықты көрсетеді деп айта алмаймыз. Өйткені тілімізде әсіресе қосымшалар арқылы жасалған сөздерде омоним құбылысы өте жиі кездеседі. Сондықтан келешекте әркез түзету, жетілдіру жұмыстарын үзбей жүргізіп отыру қажет.

ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР

1. Жұбанов А.Қ. Корпустық лингвистика – қазақ тіл білімінің жаңа бағыты // Тілтаным. Языкознание. ҚР БЖҒМ А.Байтұрсынұлы атындағы Тіл білімі институты, сәуір-мамыр-маусым. – Алматы, 2009. – №2 (34). – 3-11-бб.
2. Сүлейменова Ә.Д. Қазақ тілі үшін ұлттық корпус керек пе?// А.Байтұрсынұлы атындағы Тіл білімі институтының «Тілтаным» журналы. –2015. – №2. – 42-45-бб.
3. Жаңабекова А.Ә. Қазақ әдеби тілінің аннотацияланған мәтіндер корпусына морфологиялық білімдер жүйесін енгізудің практикалық тәжірибелері және оларды оқыту жүйесіне енгізудің ғылыми-әдістемелік негіздері // «Қазақ тіл білімі: жаңа ғылыми парадигмалар мен оқытудың инновациялық технологиялары» атты халықаралық ғылыми-теориялық конференция материалдары. – Алматы, 1912. – 197-201-бб. және т.б.
4. Жубанов А.К., Фазылжанова А. О концепции по созданию национального корпуса казахского языка (НККЯ) // Материалы Международной научно-практической конференции «Контрастивные исследования и прикладная лингвистика». – Минск, 2014.
5. Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие. – М.: Едиториал УРСС, 2003. – С. 114 с.;
6. Баранов А.Н., Плунгян В.А., Рахилина Е.В. Путеводитель по дискурсивным словам русского языка. – М., 1993.
7. Перцов Н.В. О роли корпусов в лингвистических исследованиях // Труды международной конференции «Корпусная лингвистика – 2006». – СПб.: Изд-во С.-Петербур. ун-та; Изд-во РХГА, 2006. – С. 318-331.
8. Жұбанов А. Қазақ әдеби тілінің электрондық корпусын түзудің теориялық бастаулары // Тіл және мәдениет: тілдің антропоэлектік парадигмасы. Профессор Ж.А.Манкееваның 60 жылдығына арналған республикалық ғылыми-теориялық конференцияның материалдары. – Алматы: А.Байтұрсынұлы атындағы Тіл білімі институты, 2010. – 191-197-бб.
9. Жұбанов А., Жаңабекова А. Мәтін тізбесіне мән беретін кез келді... – Алматы, «Ана тілі» газеті, 14-мамыр, 2015.
10. Құлманов С. Корпусқа енгізілетін мәтіндердегі сөздерді морфологиялық талдау және оларға шартты белгілер қою принциптері // Тіл және мәдениет: тілдің антропоэлектік парадигмасы. – Алматы: А.Байтұрсынұлы атындағы Тіл білімі институты, 2010. – 217-222 бб.
11. Жаңабекова А., Жазахметова А. Жарғылай автоматты метабелгіленімдер енгізу бағдарламасының жұмыс істеу технологиясы/ Ә.Жүнісбектің 80-жылдық мерейтойына арналған «Қазақ фонетикасы» атты дөңгелек үстел материалдары. – Алматы, 2018., 19 қазан. – 67-80 бб.
12. Жұбанов А., Жаңабекова А. Корпустық лингвистика / Монография. – Алматы: «Қазақ тілі» баспасы, 2017. – 336 б.
13. Жаңабекова А., Қаршығаева А. Мәтіндер корпусына фонетикалық белгіленімдер енгізуге арналған нұсқаулық сипаттамасы (Профессор Ә.Жүнісбектің ұшқадамдық ұстанымы бойынша) / Ә.Жүнісбектің 80-жылдық мерейтойына арналған «Қазақ фонетикасы» атты дөңгелек үстел материалдары. – Алматы, 2018., 19 қазан. – 53-58 бб.
14. Фазылжанова А. Национальный корпус казахского языка как инновационно-информационная база государственного языка (об особенностях и задачах первого этапа разработки и создания) / Alato Academic studies журналы, 2016, №3. – 103-109 бб.
15. Пірманова К. Қазақ тілі мәтіндер корпусындағы сөзжасамдық белгіленімдер әзірлемесі/ Магистрлік диссертация. – Алматы, 2018.
16. Мурсал А. Күрделі егістіктер: лексикографиялау және автоматтандыру/Магистрлік диссертация. – Алматы, 2018.
17. Мейірбекова А. Қазақ тілінің ұлттық корпусында омонимдерді ажыратудың жолдары (лингвистикалық анализ)/Магистрлік диссертация. – Алматы, 2018.
18. Жинкин Н.И. О кодовых переходах во внутренней речи // Вопросы языкознания. –1964. – №6.

19. Janice Carruthers Annotating an oral corpus using the Text Encoding Initiative. Methodology, problems, solutions¹ – Cambridge University Press Publishers, 2008.
20. Robbie Love, Claire Dembry, Andrew Hardie, Vaclav Brezina and Tony McEnery. The Spoken BNC 2014. John Benjamins Press Publishers, 2017.
21. Finegan E. Language: its structure and use. – N.Y.: Harcourt Brace College Publishers, 2004. – P. 24-25
22. Выготский Л.С. Избранные психологические исследования. Мышление и речь. Проблема психологического развития ребенка. – М., 1956. – С.519.
23. Амирова Т.А. Функциональная взаимосвязь письменного и звукового языка. – М, 1985. – С.283.
24. Амиров Р. Особенности синтаксиса казахской разговорной речи. –Алма-Ата, 1972. – С.180.
25. Күдерінова Қ. Қазақ жазуының теориялық негіздері. – Алматы, 2010.
26. Күдерінова К.Б. Ауызекі сөйлеу тілін зерттеудің маңызы // БҚМУ Хабаршысы. – 2016. – №4.
27. Львов М.Р. Основы теории речи. – М, 2002. – С.248.
28. Colleen Richey. Speech Corpora. 2020.
29. Уәли Н. Графика. Орфография. Орфоэпия. – Алматы, 2018. – 344 б.
30. Сыздық Р. Сөз сазы. Сөзді дұрыс айту нормалары. – Алматы, 1983. –75 б.
31. Балақаев М. Қазақ әдеби тілі және оның нормалары. – Алматы, 1984. – 84 б.
32. Leech, Geoffrey «The Spoken English Corpus in its context.» Foreword. Knowles, Gerard; Wichmann, Anne; Alderson, Peter, eds. (1996). Working with Speech. Longman. p. ix. - 1996. ISBN 9780582045347.
33. <https://www.culture.gouv.fr/>
34. The Lancaster Los Angeles Spoken Chinese Corpus (LLSCC) <https://www.lancaster.ac.uk/>
35. <http://www.corpusrural.es/>
36. <https://clrd.ninjal.ac.jp/cs/en/>
37. <http://russpeech.spbu.ru/>
38. <https://www.korpus.cz/>
39. <https://ruscorpora.ru/>
40. <https://qazcorpus.kz/>
41. McEnery T., Wilson A. Corpus Linguistics. – Edinburgh: Edinburgh University Press, 2001.
42. Quirk R. Svartvik J., Quirk R. A corpus of English Conversation. – Lund: Gleeup, 1980.
43. Meyer Ch. F. English Corpus Linguistics: An Introduction. Cambridge: Cambridge University Press, 2002. – Xvi + 168.
44. Чилингарян К.П. Корпусная лингвистика: теория vs методология // RUDN Journal of Language Studies, Semiotics and Semantics // 2021 Vol. 12 No. 1 196-218
45. Захаров В.П. Корпусная лингвистика: учебно-метод. пособие. – Спб., 2005.
46. Bazarbayeva Z.M. Universal Properties of Intonation Components // Review of European Studies; Vol. 7, No. 6; 2015 ISSN 1918-7173 E-ISSN 1918-7181 Published by Canadian Center of Science and Education
47. Амиева А. М., Филимонов В. В., Сергеев А. П. [и др.] // 2я Международная конференция студентов, аспирантов и молодых ученых «Информационные технологии, телекоммуникации и системы управления»: сборник докладов. – Екатеринбург: [УрФУ], 2016. – С. 251-260.
48. Современный русский язык. М., 2002. – С.46.
49. Базарбаева З.М. Қазіргі қазақ тілі интонациясының негіздері. – Алматы: Комплекс, 2002. – 202 б.
50. Кеңесбаев І.К. Известия АН КазССР. – Серия линг. – №5. – 1945.
51. Жүнісбек Ә. Қазақ тіл білімінің мәселелері. – Алматы: Абзал-ай, 2018. – 368 б.
52. Базарбаева З.М. Қазақ фонологиясының негіздері. – Алматы, 2012, 120 б.
53. Базарбаева З.М. Корпус казахского языка: просодическая разметка // Международный журнал прикладных и фундаментальных исследований. – 2016. – № 11-2. – С.334-337; URL: <https://applied-research.ru/ru/article/view?id=10495> (дата обращения: 07.04.2022).

54. Мырзабеков С. Қазақ тілінің дыбыс жүйесі. – Алматы: Сөздік-Словарь, 1999. – 201 б.
55. Пешковский А.М. Интонация и грамматика // Избранные труды. – Москва, 1959. – С.453.
56. Бондарко Л.В., Вербицкая Л.А., Гордина М.В. Основы общей фонетики. – Санкт-Петербург, 2000. – С.160.
57. Сағыбадинова М. Интонацияны зерттеудегі негізгі мәселелер мен бағыттар (ағылшын тіліндегі теориялық материал негізінде) ҚазҰУ Хабаршысы, «Филология» сериясы. №3(133) – Алматы, 2011. – 110-114 бб.
58. Сухарева Е.Е. Дифференциальные признаки завершенности и вопросительности в интонационной системе английского языка. дисс.к.ф.н. – Воронеж, 2003. – С.178.
59. Нұрмаханова Ә.Н., Жүнісбекова К. Қазақ тіліндегі сөйлемдердің интонациясын оқыту мәселелері. – Алматы, 1982 – 181б.
60. Сыздықова Р. Қазақ тілінің анықтағышы (емле, тыныс белгілері, сөз сазы). – Астана: Елорда, 2000. – 532 б.
61. Светозарова Н.Д. Интонационная система русского языка. – Ленинград: Издательство Ленинградского университета, 1982. – 173б.
62. Нұршайықов Ә. Ақиқат пен аңыз. – Алматы, 2009. – 464 б.
63. Аманбаева А.Ж., Жұмабаева Ж.Т. Ғылыми дискурстағы орфоэпиялық және просодикалық ерекшеліктер // ҚазҰУ Хабаршысы, «Филология» сериясы №3 (175). – Алматы, 2019. – 68-76 бб.
64. <https://kk.wikipedia>
65. Фазылжанова А. Сөз сазы және интонация (эксперименттік-фонетикалық зерттеу). – Алматы, 2022. – 208 б.
66. Жүнісбек Ә. Қазақ фонетикасы. – Алматы: «Арыс», 2009. – 312 б.
67. Бизақов С. Қазақ тілі қандай тіл? – Алматы: «Арыс», 2010. – 432 б.
68. Құрал О. Біз қалай сөйлейміз? – Алматы: «Асыл сөз» баспасы, 2021. – 304 б.
69. Орфоэпиялық сөздік. – Алматы: «Арыс» баспасы, 2007. – 800 б.
70. Сәбден О. Ойлы кітап. – Алматы: «Service Press», 2021. – 364 б.
71. <https://hostingkartinok.com/>
72. <https://kharshit.github.io/blog/>
73. Ұлттық картографиялық-геодезиялық қоры [Электронды ресурсы]. URL: <https://nkgf.kz/> (жарияланған уақыты: 2021)
74. Қазақстан Республикасының географиялық атауларының мемлекеттік каталогы: 14 том. – Алматы: 2012. – Т.1. – 226 б.
75. Сулейменова Ә.Д. Макросоциоллингвистика /Монография / Ә.Д.Сулейменова. – Алматы: Қазақ университеті, 2011. – 404 б.
76. University of Colorado Boulder Libraries [Электронды ресурсы] URL: <https://libguides.colorado.edu/c.php?g=456725&p=7724222>
77. Corpus of Contemporary American English [Электронды ресурсы] URL: <https://www.english-corpora.org/coca/>
78. The Turkish corpus [Электронды ресурсы] URL: <https://tscorpus.com/academic/>
79. English-Corpora.org [Электронды ресурсы] URL: <https://www.english-corpora.org/corpora.asp>
80. Фромкин В., Родман Р., Хайамс Н. Тіл біліміне кіріспе/Оқулық / В.Фромкин, Р. Родман, Н. Хайамс. – Алматы: Ұлттық аударма бюросы, 2018. – 608 б.
81. Сулейменова Ә.Д. Қазақ тілі үшін ұлттық корпус керек пе? [Мәтін] / Ә.Д. Сулейменова // ҚазҰУ хабаршысы, Филология сериясы. – 2011. №1 (131). 76-79 б.
82. Бияров Б. Жер-су аттарының сөзжасамдық үлгілері/ Монография. – Алматы: Мемлекеттік тілді дамыту институты, 2012. – 460 б.
83. Подольская Н.В. (1988) Словарь русской ономастической терминологии. – М.: Наука, – С.192.
84. Тілеубердиев Б.М. (2006) Қазақ ономастикасының лингвокогнитивтік аспектілері. – Алматы: Арыс, – 280 б.

85. Мадиева Г.Б. Имя собственное в контексте познания. – Алматы: Қазақ университеті, 2004. – С.190.
86. Бияров Б. Өр Алтайдың жер-су аттары. – Алматы, 2002. – 180 б.
87. Шайкенова А.Ж. Көркем мәтін және дискурстық талдау. Филология студенттеріне арналған оқу құралы. – Петропавл, 2012.
88. Мамонтова В.В. Корпусная лингвистика в современной языковедческой парадигме / Актуальные вопросы современной науки, 2010. –№ 12. – С. 231-235.
89. Қожахметова А. (2015) Көркем мәтін прагматикасын корпусстық лингвистика кеністігінде зерттеу/ Қазақ ҰҰ хабаршысы Филология сериясы. –№6 (158). – 300-305 бб.
90. Шаров С.А. (2003) Представительный корпус русского языка в контексте мирового опыта Научно-техническая информация – Серия 2 Информационные процессы и системы. – № 6. – С. 10.
91. Жұбанов Қ. Қазақ тілінің жоғары курсы/Лекциялар желісі. – Алматы: Қазақ университеті, 1999. –100 б.
92. Tektigul Z.O., Bayadilova-Altybayeva A.B., Sadirova K.K., Kushkarova G.K., Karabaev M.I.2 Transmissional and magical nature of naming system. A multifaceted review journal in the field of pharmacy. Systematic Review Pharmacy. – 2020. – Vol 11. – Issue 4. pp. 445-450. / <https://www.sysrevpharm.org/articles/transmissional-and-magical-nature-of-naming-system.pdf>
93. Уәли Н. Ономастикалық атаулардағы тарихи-мәдени кодтарды ашудың лингвокогнитивтік негіздері//Ономастикалық хабаршы, №2 (28) 2014. – 6-14 бб.
94. Тіл білімі терминдерінің түсіндірме сөздігі [Мәтін] / Қалиев Ғ. – Алматы: Сөздік-Словарь, 2005. – 439 б.
95. Конкашпаев Г. Казахские народные географические термины//дисс. канд. наук. – Алма-Ата, 1949.
96. Койчубаев Е. Основные типы топонимов Семиречья// дисс. канд. наук. – Алма-Ата, 1967.
97. Абдрахманов А. Некоторые вопросы топонимики Казахстана// дисс. канд. наук. – Алма-Ата, 1954.
98. Жанұзақов Т. Қазақ есімдерінің анықтамалығы. – Алматы: Арыс, 2009. – 552 б.
99. Попова В.Н. Структурно-семантическая природа топонимов Казахстана (сравнительно-историческое исследование). Автореф. дис д-ра филол наук. – Алматы, 1997.
100. Керимбаева С.Е. Ономастическая экспликация мифологической мира в казахском языке. – Алматы, 2004. – С.28.
101. Мәдиева Г. Ономастика Казахстана: аналитический обзор (подходы исследования онимов). – Алматы: Қазақ университеті, 2018. – 80 с.
102. Рысберген Қ.Қ. Қазақ топонимиясының лингвокогнитивтік және этномәдени негіздері. Филол. ғыл. доктор. дисс. авт. – Алматы, 2010. – 64 б.
103. Кенесбаев І.К. Қазақ тілінің фразеологиялық сөздігі. – Алматы, 1977. – 712 б.
104. Казахский фольклор в собрании Г.Н.Потанина. – Алма-Ата: Наука, 1972. – С.382.
105. Уәлиханов Ш. Көп томдық шығармалар жинағы. I том. – Алматы, 2010. – 376 б.
106. Мақатаев М. Таңдамалы шығармалар. /М.Мақатаев атындағы халықаралық қор. – Алматы: ҚазАқпарат, 2013. – 478 б.
107. Бөкей О. Үркер ауып барады: Повестер мен әңгімелер. – Алматы: Жалын, 1981. – 528 б.
108. Қазақ әдеби тілінің сөздігі. 12-том. – Алматы: Арыс, 2010. – 752 б.
109. Бектаев Қ. Статистико-информационная типология тюркского текста. – Алматы: Наука, 1978.
110. Жұбанов А. Основные принципы формализации содержания казахского текста. – Алматы, 2002. – 250 б.
111. Фазылжанова А. Үштілділік ұлттық мүддеге қызмет етуі тиіс //http://www.ikitap.kz. – тамыз, 2016.
112. Сичинава Д.В., Дышкант А.Н. Базы данных не книжной письменности, исторический корпус, словоуказатель: интеграция // Компьютерная лингвистика и интеллектуальные технологии: по материалам международной конференции «Диалог 2020». – Москва, 17–20 июня 2020 г.

113. Vorislamische Alttürkische Texte: Elektronisches Corpus //URL: <https://vatec2.fkidg1.uni-frankfurt.de>
114. Древнерусские берестяные грамоты. <http://gramoty.ru/>
115. Жаңабекова А.Ә., Қожахметова А.Қ. Метабелгіленім енгізілген мәтіндерді арнайы компьютерлік бағдарлама бойынша өңдеу // «Тілтаным» журналы. – №3 (83) – 2021.
116. Плунгян В.А. Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. - М., 2008. - № 2 (16). – С. 7-20 <http://www.philology.ru/linguistics2/plungyan-08.htm>
117. Сейтбекова А.А. Сейдамаг А.К. Тарихи ішкорпус әзірлеудің әлемдік тәжірибесінен // «Абылай хан атындағы ҚазХҚ және ӨТУ Хабаршысы, «Филология ғылымдары» сериясы, Том 66 № 3 (2022) <https://doi.org/10.48371/PHILS.2022.66.3.012>
118. Жаңабекова А.А., Қожахметова А.К., Бесиров Е.Б. Картотекалық қорды цифрландыру – қазақ тілінің ұлттық корпусын құрудың негізі. «Абылай хан атындағы ҚазХҚ және ӨТУ Хабаршысы, «Филология ғылымдары» сериясы Том 67 № 4 (2022) <https://doi.org/10.48371/PHILS.2022.67.4.003>
119. Жаңабекова А.Ә., Қожахметова А.Қ. Tiltanym No3 (83) 2021. – 37-52 бб. <https://www.tiltanym.kz>
120. Гаврилова Т. С., Шалганова Т. А., Ляшевская О. Н. К задаче автоматической лексико-грамматической разметки старорусского корпуса XV-XVII вв // Вестник ПСТГУ. Серия III: Филология. 2016. Вып. 2 (47). – С. 7-25.
121. Бембеев Е.В. Коллекции рукописей на старокалмыцком (ойратском) языке XVII–XIX вв. в свете компьютерной обработки: постановка проблемы// Информационные технологии и письменное наследие. E1Manuscript-2012: Материалы IV международной научной конференции (Петрозаводск, 3-8 сентября 2012 г.). – Петрозаводск, Ижевск, 2012. – С. 31-34.
122. Corpus of Historical Portuguese. <https://www.clarin.eu/resource-families/historical-corpora>.
123. Vatri A, McGillivray B. The Diorisis Ancient Greek Corpus Linguistics and Literature. Research Data Journal for the Humanities and Social Sciences. Издатель: Brill E-ISSN:2452-3666. 2018. page 55-65 /https://brill.com/view/journals/rdj/3/1/article-p55_55.xml
124. Эстер Симон. Здание корпуса из древневенгерских кодексов. In: Каталин Э. Поцелуй (ред.): Эволюция функциональной левой периферии в венгерском синтаксисе. Оксфорд: Издательство Оксфордского университета, 2014.
125. Сердюченко Г. П. Русская транскрипция для языков зарубежного востока. – Москва: Наука, 1967. – 359 с.
126. Фазылов Э.И. Староузбекский язык. Хорезмские памятники XIV века. – Т.2. – Ташкент: Фан, 1971. – С.778.
127. Иванов С. Н.Родословное древо тюрок Абу-л-Гази-хана. – Ташкент: изд.«Фан» Узбекской ССР, 1969. – С.202.
128. Әбілқасымов Ә. Әбілғазы ханның «Түркі шежіресі» және оның тілі. – Алматы: Арыс, 2001. – 246 б.
129. Сыздықова Р. Ясауи «Хикметтерінің» тілі. – Алматы, Сөздік-Словарь, 2004. – 552 б.
130. Савчук С.О. Корпус текстов XVIII века в составе национального корпуса русского языка: проблемы и перспективы. 25 июнь 2008 / <http://textualheritage.org/lt/el-manuscript-08-/xviii.html>
131. Байтұрсынов А. Ақ жол: Өлеңдер мен тәржімелер, публ. Мақалалар және әдеби зерттеу / Құраст. Р.Нұрғалиев. – Алматы: Жалын, 1991. – 464.
132. Жұбанов Қ. Қазақ тілі жөніндегі зерттеулер. – Алматы: Мемлекеттік тілді дамыту институты, 2010. – 608 б.
133. Өуезов М. Көркем аударманың кейбір теориялық мәселелері//Кітапта: Көркем аударманың кейбір мәселелері (мақалалар жинағы). – Алматы, Қазмемкөркемәдеббас, 1957. – 3-12 бб.
134. Нелюбин Л.Л. Толковый переводоведческий словарь. 3-е изд., перераб. – М.: Флинта: Наука, 2003. – С.320.

135. Виноградов В.С. Введение в переводоведение (общие и лексические вопросы). – М.: 2001. – 230 с.
136. Рецкер Я.И. Теория перевода и переводческая практика. – М.: «Международные отношения». – 1974. – С.127.
137. Бархударов Л.С. Язык и перевод (вопросы общей и частной теории перевода). – М.: «Международные отношения». – 1975. – С.342.
138. Добровольский Д.О. Лингвоспецифичная лексика в корпусах параллельных текстов // Д.О.Добровольский // Речевые жанры современного общения. Тезисы докладов междунар. конф. «11-е Шмелевские чтения (23-25 февр. 2015)». – М.: ИРЯ РАН, 2015. – С. 47-49.
139. Добровольский Д.О. Корпус параллельных текстов в исследовании культурно-специфичной лексики // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. – СПб.: Нестор-История, 2009. –С. 383-401.
140. Waldenfels R. Compiling a parallel corpus of Slavic languages. Text strategies, tools and the question of lemmatization in alignment // Beitrage der Europaischen Slavistischen Linguistik (POLYSLAV). 9. Munchen, 2006. –S. 123-138.
141. Сичинава Д.В. Параллельные тексты в составе национального корпуса русского языка: новые направления развития и результаты // Труды Института русского языка РАН. – М., 2015. – С.194-234.
142. Тао Ю., Захаров В.П. Разработка и использование параллельного корпуса русского и китайского языков // НТИ. Сер. 2. Информ. Процессы и системы. 2015. №4. – С. 18-27.
143. Englund-Dimitrova V. Tolken's roll: perspektiv från nyare forskning ochimplikationer för tolkutbildning // Материалы Научно-практического семинара по преподаванию устного перевода (шведский язык/русский язык). – М., 1999. – С. 36-47.
144. Бельгер Г. Жизнь – эпопея (эссе-триптих, статьи). – Алматы: ИД «Жибек жолы», 2009. – С.144.
145. Жаксылыков А.Ж. Особенности нового перевода романа-эпопеи М.О. Ауэзова «Путь Абая» на русский язык // Художественный перевод и литературный процесс (избранные лекции и исследования). – Алматы, 2013. – С. 167-178.
146. Болатова Г.Ж. Жаңа «Абай жолы»: А. Кимнің аудармасындағы кейбір ерекшеліктер // ҚазҰУ Хабаршысы. Филология сериясы. №2 (166). 2017. – 320-324-бб.
147. Жанпейісов Е. М.Әуезовтің «Абай жолы» эпопеясының тілі. – Алматы: «Ғылым» баспасы, 1976. – 168 б.
148. Момынова Б.К. Қазақ ым-ишараттары: сөйлеу этикеті мен қатысым мәдениетінде. – Алматы: «Арыс» баспасы, 2016. – 496 б.
149. Әуезов М. Абай жолы. I том. – Алматы: «Жазушы», – 1989. – 377 б.
150. Амирова Ж.Г. Аудармашы қызметінің кәсіби негіздері/Әдістемелік құрал. – Алматы: «Тіл» оқу-әдістемелік орталығы, 2009. – 144 б.
151. Ауэзов М. Путь Абая. Роман, т.І. – Алма-Ата, «Жазушы», 1977. – С.608.
152. Veronis J. From the Rosetta Stone to the Information Society: A Survey of Parallel Text Processing // J. Veronis (ed.). Parallel Text Processing. Kluwer, 2000. P. 1-25.
153. King P. Trialling a Multilingual Parallel Concordancer // Second International Conference on Current Trends in Studies of Translation and Interpreting. Abstracts. Hungary, 1996. – P. 49-50.
154. Langlois L. Bilingual Concordances: A New Tool for Bilingual Lexicographers // Expanding MT Horizons. Proceedings of the Second Conference of the Association for Machine Translation in the Americas. Montreal, Quebec, Canada, 1996. – P. 34-42.
155. Li H.F., Heo N.W., Moon K.H., Lee J.H., Lee G.B. Lexical Transfer Ambiguity Resolution Using Automatically-Extracted Concept Co-occurrence Information // International Journal of Computer Processing of Oriental Languages. World Scientific Publishing Co, 2000. № 13(1). – P. 53-68.
156. Hong C.S., Pak M.G. Developing a Large Scale Computational Lexical Database of Contemporary Korean: SELK // Proceedings of the 19th International Conference on Computer Processing of Oriental Languages. Seoul, Korea, 2001. – P. 20-26.

157. Dan Melamed I. *Empirical Methods for Exploiting Parallel Texts*. MIT Press, Cambridge (MA), 2001.
158. Жұбанов А. Компьютерлік лингвистикаға кіріспе. – Алматы: «Арыс» баспасы, 2008. – 640 б.
159. Hinkel E. *Second language writers' text* (2002). – Mahwah, NJ, 2002. – 392 p.
160. Granger S. (ed). *Learner English on computer* (1990). London: Addison Wesley Longman. – 1990. – 250 p.
161. Granger S. *The computer learner Corpus: a versatile new source of data for SLA research* (1998) // *Learner English on Computer*. L., 1998. – 16 p.
162. Мальцева М.С. «Учебный корпус как база для лингвистического и лингводидактического анализа в рамках методики преподавания иностранных языков» (2011) // *Социально-экономические явления и процессы*, № 9 (031), 2011. – С. 209-212
163. Соснина Е.П. К анализу лингвистических ошибок учащихся на базе учебного корпуса текстов (2017). – 2017. – С. 39-44.
164. Грудева Е.В. Корпусы ошибок: целевая аудитория, возможная архитектура корпуса (2018). – 2018. – С. 63-72.
165. Granger S. (ed). *Learner English on computer*. London: Addison Wesley Longman, 2010. – 512.
166. Mauranen Anna and Elina Ranta. *English as a Lingua Franca: Studies and Findings*. – Newcastle: Cambridge and Scholars, 2009.
167. Камшилова О. Н. Иностранные языки в дистанционном обучении: мат-лы III Междунар. науч.-практ. конф. Т. 2 (Пермь, 23-25 апр. 2009 г.). URL: <http://window.edu.ru>.
168. Соснина Е.П. Прикладная Лингвистика. О разработке и использовании российского учебного корпуса переводов. URL: <http://ling.ulstu.ru>. 2019. – 242.
169. Рыков В. В. Корпус текстов и речевая деятельность - проблемы подобия // *Корпусная лингвистика -2006: тр. междунар. конф. 10-14 окт.* – Санкт-Петербург, СПб., 2006. – С. 347-355.
170. Жұмабекова А. Қазақ тілінің антонимдер сөздігі. – Алматы, 2000. – 128 б.
171. <https://sozdikqor.kz/>
172. Қазақ тілінің диалектологиялық сөздігі. I-II тт. – Алматы, 1996-1999.
173. Қазақ тілінің аймақтық сөздігі. – Алматы, 2005.
174. Сарыбаев Ш. Диалектологический атлас казахского языка // *Советская тюркология*, 1972, №3, С. 85; Айдаров Т. Лингвистикалық география. – Алматы, 1977 ж.; Соныкі. Проблемы диалектной лексикологии и лингвистической географии. – Алма-Ата, 1991; Нақисбеков О. Интерпретация некоторых фонетических явлений по материалам диалектологического атласа казахского языка // VII региональная конференция по диалектологии тюркских языков. – Алма-Ата, 1973. – С. 82.
175. Омарбеков С., Нақисбеков О. Қазақ тілінің лингвистикалық атласы жөнінде // *Тілтанымдық зерттеулер* (2000-2002 жж.). – Алматы, 2007.
176. Базылхан Б. Моңғолиядағы қазақтардың тілі. – Ұланбатыр, 1991.
177. Үдербаев А. Моңғолиядағы қазақ диаспорасы тіліндегі фонетикалық ареал құбылыстарды салыстырмалы зерттеу. Канд. дисс. – Алматы, 2002; Б.Бәмішұлы, Моңғолиядағы қазақтардың тілінің жағдайы. – Алматы, 2004.
178. Мұстафаұлы С. Тұрақты тіркестердегі жергілікті ерекшеліктер (Қытайдағы қазақтар тілі материалы бойынша). – Алматы, 2003.
179. Өтебеков Б. Қазақ диаспорасының тілі (Ауғанстан, Иран қазақтарының тілінен жиналған материал бойынша). – Алматы, 2000.
180. Качинская И.Б. Корпус диалектных текстов в Национальном корпусе русского языка: состояние и перспективы // *Лексический атлас русских народных говоров* (Материалы и исследования). 2009. СПб. – С. 57-68.
181. Качинская И.Б. 2011. Диалектный Подкорпус НКРЯ. Новый стандарт подачи. Новое рабочее место // В сб.: *Русская устная речь. Материалы международной научной конференции*

«Баранниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура общения». Межвузовское совещание «Проблемы создания и использования диалектных корпусов». Саратов, СГУ, 15-17 ноября 2010 г., ред. О.Ю. Крючкова, А.И.Буранова, В.Е.Гольдин, Л.В. Балашова. – С. 245-255.

182. Летучий А. Б. 2005. Корпус диалектных текстов: задачи и проблемы // Национальный корпус русского языка: 2003-2005. М.: Индрик. – С. 215-232.

183. Летучий А. Б. 2009. Диалектный корпус: состав и особенности разметки // Национальный корпус русского языка: 2006-2008. Новые результаты и перспективы. – СПб.: Нестор-История. – С. 114-128.

184. Русская устная речь. Материалы международной научной конференции «Баранниковские чтения. Устная речь: русская диалектная и разговорно-просторечная культура общения». Межвузовское совещание «Проблемы создания и использования диалектных корпусов». – Саратов, СГУ, 15-17 ноября 2010 г., ред. О. Ю. Крючкова, А. И.Буранова, В.Е.Гольдин, Л.В. Балашова. Юрина Е. А. (2011), Томский диалектный корпус: в начале пути // Вестник Томского гос. ун-та. Филология. № 2. – С. 58-63

185. Крючкова О. Ю. Диалектологический корпус как источник лингвокультурологического изучения русских народных говоров. - Саратов <https://cyberleninka.ru/article/n/dialektologicheskii-korpus-kak-istochnik-lingvokulturologicheskogo-izucheniya-russkih-narodnyh-govorov>

186. Свешникова Н. В. История села в народном слове (с. Белогорное Вольского р-на Саратовской области) // Культурологические и лингвистические аспекты коммуникации. – Саратов, 2008. Вып. 8.

187. Юрина Е.А. Томский диалектный корпус: в начале пути //Вестн. Том. гос. ун-та. Филология. – 2011. – №2 (14). – С. 58-63.

188. Пурицкая Е.В. 2012. Диалектный подкорпус национального корпуса русского языка как источник изучения лексической динамики диалекта // Северно-русские говоры. Вып. 12: межвуз. сб. / отв. ред. А.С. Герд. – СПб.: Изд-во С.-Петерб. ун-та, 2012. – С. 14-22.

189. Сичинава Д.В., Качинская И.Б. 2014. Корпус диалектных тестов в национальном корпусе русского языка: сегодняшнее состояние и перспективы // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4-8 июня 2014 г.). Вып. 13 (20). – М.: Изд-во РГГУ, 2014. – С. 593-600.

190. Крючкова О.Ю., Гольдин В.Е. Корпус русской диалектной речи: концепция и параметры оценки. – Саратов, <https://docplayer.com/57503155-Korpus-russkoy-dialektnoy-rechi-konceptsiya-i-parametry-oc>

191. Исабеков С. Принцип дополнительности в номинативной системе языка (на материале немецкого и казахского языков: автореф. докт.дисс. – Алматы, 1996.

192. Авакова Р.А. Фразеосемантика. – Алматы: 2013.

193. Уэли Н. Фразеология және тілдік норма. – Алматы, 1998.;

194. Древнетюркский словарь. – Л., 1969.

195. Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. – Алматы, 2017. – II том.

196. Маслова В.А. Лингвокультурология. – М.: 2004.

197. Қайдар Ә. Қазақтар ана тілі әлемінде: этнолингвистикалық сөздік. – Алматы: Дайк-Пресс, 2009. – 780 б.

198. Золотова Г.А. Аспекты изучения категории глагольного времени // Известия Академии наук СССР. – Серия литературы и языка. 1975. Т. 34, №3. – С. 248-258.

199. Балақаев М. Қазақ тіл білімінің мәселелері. – Алматы: Арыс, 2008. – 592 б.

200. Уэли Н. Қазақ сөз мәдениетінің теориялық негіздері /Фил. ғыл. докторы ғылыми дәрежесін алу үшін дайындаған диссертациясының авторефераты. – Алматы, 2007.

201. Жұбанов А.Қ. Қазақ тілі мәтіндер корпусының компьютерлік базасын құрудың алғышарттары // Әдеби тіл және қазақ тілінің өміршеңдігі / Жауапты ред. Қ.Күдерінова. – Алматы: «КИЕ» лингвоелтану инновациялық орталығы, 2009. –175-179 бб.

202. Сулейменова Э.Д. Қазақ тілі үшін ұлттық корпус керек пе? // ҚазҰУ Хабаршысы. Филология сериясы, № 1 (131). – 76-79 с.
203. Жұбанов А.К., Фазылжанова А. О концепции по созданию национального корпуса казахского языка (НККЯ) // Материалы Международной научно-практической конференции «Контрастные исследования и прикладная лингвистика». – Минск, 2014.
204. Жаңабекова А.Ә. Қазақ тілі мәтіндеріне морфологиялық белгіленім коюдың ғылыми-тәжірибелік негіздері // Научно-практическая конференция «Языки меньшинств в компьютерных технологиях: опыт, задачи и перспективы». – Уфа, 2012. – С. 42-50.
205. Жұбанов А.К. Принципы автоматизации морфологической разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы международной научно-практической конференции «Контрастные исследования и прикладная лингвистика». – Минск: МГЛУ, 2015. – С. 111-119.
206. Койбагаров К.Ч., Мусабаев Р.Р., Кулманов С.К. Разработка алгоритмов автоматического анализа словоформ казахского языка // «Қазіргі қазақ тіл білімі: қолданбалы лингвистиканың өзекті мәселелері»: халықаралық ғылыми-теориялық конференция материалдарының жинағы. – Алматы, А.Байтұрсынұлы атындағы Тіл білімі институты, 2012. – С. 83-87.
207. Жаңабекова А., Пирманова К. Технология программы полуавтоматической метаразметки казахского национального корпуса // Актуальные проблемы современной лингвистики и гуманитарных наук: сборник статей XI Всероссийской научно-методической конференции с международным участием. – М.: РУДН, 2019. – С. 465-474.
208. Уәли Н. Ұлттық мәдениеттің мәдени кодтарын ашу мәселелері/ А.Байтұрсынұлы атындағы Тіл білімі институтының «Тілтаным» журналы. –2015. №3 – 40-52 бб.; Уәлиев Н. Ай аттары асылымыз. Күн қайыру // Ана тілі, 1991. – №5, №8, №10, №12.
209. Қайдар Ә. Қазақтар ана тілі әлемінде: этнолингвистикалық сөздік. – Алматы: Дайк-Пресс, 2009. – 780 б.
210. Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. Том 1-5. – Алматы: «Әлем. Даму. Интеграция», 2017.
211. Байтұрсынұлы А. Алты томдық шығармалар жинағы. Т.1. – Алматы: Ел-шежіре, 2013. – 384 б.
212. Гумбольд В. Избранные труды по языкознанию. – М.: Прогресс, 2000. – С. 400.
213. Сепир Е. Избранные труды по языкознанию и культурологии. – М.: Издательская группа «Прогресс», «Универс», 1993. – С.656.
214. Поршнев В.Ф. О начале человеческой истории. – М.: Мысль, 1974. – 487 с.
215. Карасик В.В. Языковая кристаллизация смысла. – М.: Гнозис, 2010. – С.428.
216. Верещагин Е.М., Костомаров В.Г. Язык и культура. – Москва, Берлин: Директ-Медиа, 2014; Алефиренко Н.Ф. Лингвокультурология: ценностно-смысловое пространство языка. – М.: Флинта, 2020.
217. Қасиманов С. Қазақ халқының қолөнері. – Алматы, Қазақстан, 1969.
218. Қайдаров А.Т. Структура односложных корней и основ в казахском языке. – Алма-Ата: Наука, 1986. – С.328.
219. Қайырбаева Қ.Т. Қазақ тіліндегі этномәдени атаулардың символдық мәні: филол. ғыл. канд. дисс. – Алматы, 2004. – 127 б.
220. Бабалар сөзі. – Астана: Фолиант, 2013. – 432 б.
221. Уәли Н. «Қамшы» концептісі // «Акад. Ш.Ш. Сарыбаев және қазақ тіл білімі мәселелері» атты ғылыми-теориялық конференция материалдары. – Алматы: Дайк-пресс, 2005.
222. Finegan E. Language: its structure and use. – N.Y.: Harcourt Brace College Publishers. 2004.
223. Жұбанов А.К. Принципы автоматизации морфологияческой разметки текстов Национального корпуса казахского языка (НККЯ) // Материалы Международной научно-практической конференции «контрастные исследования и прикладная лингвистика» 29-30 октября 2014 года в Минском государственном лингвистическом университете. – Минск. – 4 комп.стр.
224. Жұбанов А.Қ. Компьютер қазақ сөзін жатсынбайды. – Алматы: «Ана тілі», №2 (34). – 3-11 б.

225. Жұбанов Қ. Қазақ тілі жөніндегі зерттеулер. – Алматы: Ғылым, 1999. 325 б.
226. Телия В.Н. О специфика отображения мира психики и знания в языке. // *Сущность, развития и функция языка*. – М., 1987. – С.28.
227. Қайдаров Ә.Т. Қазақ тілінің өзекті мәселелері. – Алматы: «Ана тілі», 1998. – 308 б.
228. Емельянова М.В. Гипертекстовый мультимедийный словарь лингвокультурного дискурса / М. В. Емельянова, Е.Н. Трегубова // *Русский язык: исторические судьбы и современность: II Международный конгресс исследователей русского языка (Москва, МГУ им. М. В. Ломоносова, филологический факультет, 18 – 21 марта 2004 г.): Труды и материалы / Составители М. Л. Ремнёва, О. В. Дедова, А. А. Поликарпов*. – М.: Изд-во Моск. ун-та, 2004. – С.472-473.
229. Трегубова Е.Н. Лингвокультурный региональный дискурс и современные способы его лексикографического описания / Е.Н. Трегубова, М.В. Емельянова // *Конференция грантодержателей регионального конкурса Российского гуманитарного научного фонда и администрации Краснодарского края «Северный Кавказ: традиции и современность. Сборник тезисов*. – Краснодар: Просвещение-Юг, 2007.
230. Ғұн тарихының жылнамасы (аударған М.Алпамысұлы). – Түркістан, 2004. – 330 б.
231. Ғабитханұлы Қ. Қытай жылнамаларындағы түркілік ономастика мәселелері // «Қазіргі заманғы түркология ғылымының өзекті мәселелері және алдағы міндеттері» атты халықаралық конгресс материалдары. – Түркістан, 2003. – 117-121 бб.
232. Қоңырбай Ә. «Көне дәуір күмбірі» (Қытай тарихындағы қазаққа қатысты деректер) // *Жалын*, 1999. – №5. – 22-25 бб.
233. Аманжолов Қ., Рахметов Қ. Түркі халықтарының тарихы. – Алматы: Білім, 1996. – 304 б.
234. Таным тармақтары: Қытай тарихшыларының көне тарихымыз жөніндегі пайымдаулары (құрастырған Қ.Салғараұлы). – Алматы: Санат, 1998. – 280 б.
235. Байжанов Т. Қазақ тіліндегі әскери лексика. Филол. ғыл. докт. ... дисс. автореф.: 10.02.02. – Алматы, 1991. – 43 б.
236. Барлыбаев С. Қазіргі қазақ тіліндегі қоғамдық-саяси лексика. – Алматы: Мектеп, 1978. – 144 б.
237. Баскаков Н.А. К этимологии собственных тюркских имен в русских летописях // *Теория и практика этимологических исследований*. – М., 1985. – 104 с.
238. Кляшторный С.Г. Древнетюркские рунические памятники как источник по истории Средней Азии. – М., 1964. – С.353.
239. Бернштам А.Н. Социально-экономический строй Енисейских тюрков. – М.,-Л., 1946. – С.211.
240. Жанұзақов Т. Қазақ есімдерінің тарихы. – Алматы: Ғылым, 1971. – 217 б.
241. Григорьев В.В. О скифских народе саках. – СПб., 1871. – 205 с.
242. Сыздыкова Р.Ғ. XVIII-XIX ғасырлардағы қазақ әдеби тілінің тарихы. – Алматы: Ғылым, 1984. – 256 б.
243. Қайдаров Ә. Тарихи лексикология және этнолингвистика. – Кітапта: Қазақ тілі тарихи лексикологиясының мәселелері. – Алматы, 1988. – 33-38-бб.;
244. Сыздыкова Р. Сөздер сөйлейді. – Алматы, 1980, 1994, 2004.
245. Жанпейісов Е. Этнолингвистика // *Ана тілі*, 1994, № 3, 20 қаңтар, Этнокультурная лексика казахского языка (на материалах произведения М.О. Ауэзова). – Алматы, 1989.
246. Уалиев Н. Сөз мәдениеті (1984), Қазақтың дәстүрлі сөзсаптамасы. – Алматы, 1997.
247. Манкеева Ж. Қазақ тіліндегі этномәдени атаулардың танымдық негіздері. – Алматы, 2008.
248. Шойбеков Р. Қазақ зергерлік өнерінің лексикасы. – Алматы, 1992.
249. Құмарұлы Ш. Қазақтың киіз үй мәдениеті. Шынжаң әсемөнер-фотосурет баспасы. – Үрімжі, 2009. – 84 б.
250. Құмарұлы Я. Қазақтың салт-дәстүрлері. Шынжаң халық баспасы. – Үрімжі, 2006.
251. Қианятұлы Н. Қазақтың дәстүрлі мал шаруашылығы мен мал шипагерлігі. – Үрімжі, Шынжаң ғылым-техника баспасы, 2003.

252. Кустова Г.И. Семантические переходы в абстрактных существительных // Вестник Новосибирского государственного педагогического университета, 2016. – Т. 6. – № 2. – С. 73-85.
253. Казачук И.Г., Л.П.Юздова Понятие «валентность» как объект лингвистики и методики обучения русскому языку в вузе // Вестник Южно-Уральского государственного гуманитарно-педагогического университета №7. – 2019. – С.125-135.
254. Апресян Ю. Д. Избранные труды. – М.: Школа «Языки русской культуры», 1995. – Т. I: Лексическая семантика (синонимические средства языка). – С.364.
255. Willems, K. Logical polysemy and variable verb valency // Language Sciences. – 2006. - No 28. – P. 580-603
256. <http://www.natcorp.ox.ac.uk/>
257. Захаров В. П. Экстралингвистическая разметка. Мегаданные [Электронный ресурс] // URL: http://download.yandex.ru/class/zakharov/CL_L4.ppt (дата обращения: 17.08.2023).
258. Караулов Ю.Н. Показатели национального менталитета в ассоциативно вербальной сети // Языковое сознание и образ мира. – М., 2000. – С.129.
259. <http://tbikartoteka.kz/>
260. Саблин М. В., Гирия Е. Ю. Артефакт из Ливенцовки – Свидетельство присутствия человека на территории Восточной Европы в интервале 2,1 – 1,97 млн лет назад // Древнейшие миграции человека в Евразии. Материалы международного симпозиума. – Новосибирск, Изд. ИАЭ СО РАН, 2009. –С.166-174.
261. Amirbekova, A.B., Esenova, K.U., Tymbolova, A.O. Lingocognitive mechanisms of formation of national world-views of Kazakh people // Middle East Journal of Scientific Research, 2013, 14(6). – С.779-783
262. Baimurzina, A.A., Amirbekova, A.B., Khabiyeva, A.A., Abitova, G.O. The role of the Russian language in the determination of Kazakh students | El papel de la lengua Rusa en la determinación de los estudiantes Kazajos // Opcion, 2019, 35(Special Issue 23), P. 800-817.
263. Rykov V.V. Corpus of texts as an implementation of the object-oriented paradigm // Proceedings of the International Seminar Dialogue-2002. – М.: Nauka, 2002.
264. Baranov A.N. Corpus linguistics // Baranov A.N. Introduction to Applied Linguistics: Study Guide. – М.: Editorial URSS, 2003. – P. 114.
265. Baranov A.N. Introduction to Applied Linguistics. – М., 2007.
266. Zakharov V.P. Corpus linguistics: Study methodical allowance. – Spb., 2005.
267. Grishina E.A., Savchuk S.O. The national corpus of the Russian language as a tool for studying the variability of grammatical norms // Proceedings of the international conference «Corpus linguistics – 2008» October 6-10, 2008 – SPb., 2008. – P. 161-169.
268. Zubov A.V., Zubova I.I. Information technology in linguistics: Textbook. allowance. – М.: Publishing Center «Academy», 2004. – 208 p.
269. Belyaeva L.N. Lexicographic potential of the parallel corpus of texts // Proceedings of the international conference «Corpus linguistics – 2004». – SPb., 2004. – P. 55-64.
270. Mitrofanova O.A., Grachkova M.A., Shimorina A.S. Automatic classification of vocabulary in parallel texts (based on texts from the Russian-Slovak corpus of parallel texts PARUS) // Proceedings of the V International Scientific and Practical Conference «Applied Linguistics in Science and Education: Linguistic Technologies and Innovative Educational Environment». – SPb.: «LEMA», 2010. – P. 231-235.
271. Mordovin A. Yu. On the issue of genre usefulness of modern non-specialized text corpora // Bulletin of ISLU, 2009. – № 2. – P. 48-52.
272. Қайдар Ә. Халық даналығы. – Алматы: Толғанай Т, 2004. – 560 б.
273. Адамбаев Б. Мақал-мәтел – халық даналығы. Қазақ халық әдебиеті. Көп томдық. – Алматы: Ғылым, 1996. – 5-6 бб.
274. Lic. Daymette Montenegro Morales. THE HISTORY OF ADVERTISING. // CD de Monografías 2012 (c) 2012, Universidad de Matanzas “Camilo Cienfuegos”.
275. Ахмет Байтұрсынұлы «Тіл-құрал» басылымдарының графикалық проекциясы (Дыбыс жүйесі мен түрлері, 1-кітап. 1914, 1924, 1927) / Ғылыми түсініктемесін жазып, құрастырғандар:

- Е.Маралбек, Г.Мамырбек, Қ.Ғабитханұлы. Жоба жетекшісі А.Фазылжан. – Алматы: ЖК «Асыл», 2023. –321 б.
276. Момынова Б. «Қазақ» газеті (1913-1918): әліпби, емле ерекшеліктері және лексика-грамматикасы. – Алматы: Қазақ тілі, 2022. – 256 б.
277. Тұрғараева Г. Ахметтің тағдыр толқыны. – Алматы: «Қазақ университеті», 2021. – 317 б.
278. Байдуллаева А. Құрылымы әртүрлі тілдердегі жарнама слоганы. Философия докторы (PhD) ғылыми дәрежесін алу үшін дайындалған диссертация. – Алматы, 2016.
279. Увайсова М.Жарнама тілі: лингвопрагматикалық аспект. Филология ғылымдарының кандидаты ғылыми дәрежесін алу үшін дайындалған диссертация. – Алматы, 2010.
280. Маломошнова А.В., Мусинова Г.С. Жарнамалық мәтіндердегі гендерлік бағдарланған лексика. Л.Н.Гумилев атындағы ЕҰУ Хабаршысы, Филология сериясы. №1, 2023.
281. Шындалиева М. БАҚ-тағы жарнама технологиясы (оқу құралы). -Астана: Өркен, 2012. – 134 б.
282. Лабов У. Логика нестандартного английского языка / В книге П. Джильоли (ред.) Язык и социальные Контекст. – Хармондсворт: Пингвин, 1972. – С. 179-215.
283. Хохлова Н.В. Корпус как инструмент социолингвистики/ Вестник гуманитарного факультета ИГХТУ. – 2014. – Выпуск 6. – С. 188-190.
284. https://adilet.zan.kz/kaz/docs/Z970000151_#z26
285. Қазақ әдеби тілінің сөздігі. Он бес томдық. 6 - том / Құраст.: Ж. Қоныратбаева, Ғ.Қалиев, Қ.Есенова және т.б. – Алматы, 2011. – 752 б.
286. Булатова Э.В. Стилистика текстов рекламного дискурса/ учеб. пособие.– Екатеринбург: Изд-во Урал. ун-та, 2012. – С.264.
287. Розенталь Д.Э., Кохтев Н.Н. Язык рекламных текстов. – М., 1981.
288. Кохтев Н.Н. Стилистика рекламы. – М., 1991.
289. Ученова В.В., Старых Н.В. История рекламы. – СПб., 2002.
290. Фещенко Л.Г. Структура рекламного текста. – СПб., 2003
291. Иншакова Н.Г. Особенности редактирование рекламных текстов // Вестн. Моск. Ун-та. Сер. 10. Журналистика. 2011. № 2. – С. 28-37.
292. «Жарнама туралы» Қазақстан Республикасының 2003 жылғы 19 желтоқсандағы N 508 Заңы.
https://adilet.zan.kz/kaz/docs/Z030000508_
293. Асанбаева С.А. Жарнама саласындағы мемлекеттік тілді қолданудың социолингвистикалық өзекті мәселелері: фил.ғыл.канд. дисс. – Алматы, 1999. – 132 б.
294. Солтанбекова Ғ.Ә. Қазақ жарнамалары: тілдік сипаты және ықпал ету қызметі: фил. ғыл.канд. дисс. – Алматы, 2001. – 138 б.
295. Беклешов Д.В. Реклама в торговле. – М.: Междунар. отношения, 1968. – С.164.
296. Пронькина А.В. Культурологический анализ национальных моделей массовой культуры США и России: дисс. ... к.к.н. – Саранск, 2008. – С.185.
297. Ежова Е.Н. Медиа-рекламная картина мира: люди и вещи. – М.: Илекса, 2010. – С.341.
298. Авдеева З.К., Коврига С.В., Макаренко Д.И., Максимов В.И. Когнитивный подход в управлении // ControlSciences. – 2007. – №3. – С.87.
299. Зағыпаров Қ.Ф. Философиялық терминдердің сөздігі. – Павлодар: ПМПИ, 2011. – С.68.
300. Баранов А.Н. Лингвистическая теория аргументации (когнитивный подход): Автореф. дисс.докт.филол.наук. – М., 1990. – С.9.
301. Кифер Ф. О пресупозициях // Новое в зарубежной лингвистике. Лингвистика текста. – М., 1978. – Вып.8. – С. 337-369.
302. Буряков Е. Эпидемия убеждений // Рекламодатель: теория и практика. – 2004. - №2. – С.86-88.
303. Баранов А.Н., Добровольский Д.О. Англо-русский словарь по лингвистике и семиотике. – М., 1996. – С.647.

304. Драздова Т.В. Научный текст и проблемы его понимания (на материале научных экономических текстов): Дисс.докт.филол.наук. – М., 2003. – С. 52-58.
305. Дудина Е.П. Семантика и функции лексико-грамматических единиц в рекламе: Автореф. дисс.канд.филол.наук. – Архангельск, 2006. – 6 с.
306. Караулов Ю.Н. Русский язык и языковая личность. – М., 1987. – С.256-260.
307. Караулов Ю.Н. Способы существования элементарных единиц знания в обыденном языковом сознании // Язык и действительность: Сб. науч.тр.памяти В.Г.Гака. – М.: Ленанд, 2007. – С. 53-61.
308. Грайс Г.П. Логика и речевое общение // Новое в зарубежной лингвистике. – Вып. 16: Лингвистическая прагматика. – М.: Прогресс, 1985. – С. 217-237.
309. Байтұрсынов А. Тіл тағылымы (қазақ тілі мен оқу-ағартуға қатысты еңбектері). – Алматы, 1992. – 448 б.
310. Ысқақов А. Қазіргі қазақ тілі: жоғарғы оқу орындарының филология факультеттеріне арналған оқулық. – Алматы: «Мектеп» баспасы, 1974. – 407 б.
311. Қазақ грамматикасы. Фонетика, сөзжасам, морфология, синтаксис. – Астана, 2002. – 784 б.
312. Нургатина А. Р. Абстрактные имена существительные в татарском языке: автореф. дисс. канд. филол. наук. – Казань, 2012. – С.27.
313. Кубрякова Е.С. Когнитивный взгляд на природу частей речи // Словарь. Грамматика: сб. ст. – М.: Ин-т рус. языка им. В.В.Виноградова, 1996. – С. 218-224.
314. Оралбай Н., Құрманәлиев К., Балтабаева Ж. Қазақ тілінің сөзжасамы: оқулық. – Алматы: Инжу-Маржан баспаханасы, 2014. – 290 б.
315. Байтұрсынов А. Ақ жол: Өлеңдер мен тәржімелер, публ. Мақалалар және әдеби зерттеу / Құраст. Р.Нұрғалиев. – Алматы: Жалын, 1991. – 464 бет.
316. Сүлейменова Э.Д. Қазақ тілі үшін ұлттық корпус керек пе? // ҚазҰУ хабаршысы. Филология сериясы. – №1(131). 2011. – 76-79 бб.
317. Қазақ әдеби тілінің сөздігі. 2 т. – Алматы, 2011.
318. Қазақ дәстүрлі мәдениетінің энциклопедиялық сөздігі. – Алматы: «Сөздік-Словарь», 1997. – 368 б.
319. Жарықбаев Қ.Б. Жантану атауларының түсіндірме сөздігі. – Алматы: Сөздік словарь баспасы, 2006. – 387 б.
320. Оразов М. Қазақ тілінің семантикасы. – Алматы, Рауан, 1991. – 216 б.
321. Finegan E. Language: its structure and use. N. Y.: Harcourt Brace College Publishers, 2004.
322. McEnery T., Wilson A. Corpus Linguistics. – Edinburgh, 2001.
323. Баранов А.Н. Корпусная лингвистика. Введение в прикладную лингвистику: Учебное пособие. – М., 2003. – С.114.
324. Зубов А.В., Зубова И.И. Информационные технологии в лингвистике: Учеб. Пособие. – М., 2004. – С.208.
325. Сиразитдинов З.А. Моделирование грамматики башкирского языка. – Уфа, 2006.
326. Meyer Ch. F. English Corpus Linguistics & An Introduction. Cambridge, 2002. – p.168.
327. Kyo M., Rissanen M. A Language in transition: the Helsinki Corpus of English texts, ICAME Journal, 1992.
328. Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. – М., 2005. – С. 155-174
329. Куканова В. Принципы семантической разметки Национального корпуса Калмыцкого языка // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. – М., 2005. – С. 187-192.
330. Исаев С. Қазіргі қазақ тіліндегі сөздердің грамматикалық сипаты. – Алматы, 1998. – 304 б.
331. Момынова Б. Қазақ тілінің морфологиясы. – Алматы, 2012. – 239 б.

332. Оразов М. Қазақ тілінің семантикасы. – Алматы, 1991. – 216 б.
333. Рыков В.В. Корпус текстов как реализация объектно-ориентированной парадигмы // Труды Международного семинара Диалог-2002. – М.: Наука, 2002.
334. Маманов Ы. Қазақ тіл білімінің мәселелері. – Алматы: «Абзал-Ай» баспасы, 2014. – 640 б.
335. Хасенова А. Етістіктің лексика-грамматикалық сипаты. – Алматы: Ғылым, 1971. – 307 б.
336. Қазақ тілі грамматикасы. Морфология. – Алматы, 1967. – 647 б.
337. Қазіргі қазақ тілі. – Алматы, 1962. – 510 б.
338. Қайдаров Ә. Қазақ тіліндегі қос сөздер: зерттеу және сөздік. – Алматы, 2013.

ГЛОССАРИЙ

Абсолютті балама – аударылатын тілдегі сөз түпнұсқа тілдегі сөзге мағыналық жағынан да, грамматикалық құрылымы жағынан да сәйкес балама.

Авторизацияланған аударма – шығыс мәтінді аудармашы түпнұсқаның авторымен бірлесе отырып аударған және автор оның жариялануын мақұлдаған аударма түрі.

Авторлық аударма – шығыс тілдегі мәтіннің авторы болып табылатын аудармашының өзі аударған аударма.

Авторлық емес аударма – шығыс тілдегі мәтіннің авторы болып табылмайтын аудармашы жасаған аударма.

Аймақтық сөздік – ұлт тілі сөйленістерінің лексикалық ерекшеліктерін қамтитын сөздік.

Айтылым – 1. Коммуникация теориясында адамдар арасындағы тілдік қатынасты жүзеге асыратын сөйлесім әрекетінің түрі, тілдік қарым-қатынас барысында адамның өз ойын жарыққа шығару үдерісі, өз сөзін екінші біреуге ұғынықты етіп жеткізуі. 2 Фонетикада <грек. akustikos – естілетін, тыңдалатын> дыбыстау мүшелері арқылы дыбыстың пайда болуын зерттейтін ілім.

Ақпаратты іздеу базасы (немесе индекс) – қажетті лексикалық бірлік туралы, сондай-ақ құжат, сайт туралы ақпарат салынған арнайы компьютерлік құрылым.

Аннотациялау – қазақ әдеби тілінен сұрыпталып алынған таңдама мәтіндердегі сөзформаларға лингвистикалық мәліметтер беріп ақпараттандыру.

Антроним – адамдардың өзіне меншікті есімдері, аты-жөні, жалған, бүркеншік, жасырын аттары, т.б.

Арнайы корпус – қандай да бір зерттеушілік мақсатты шешу үшін құрастырылған көлемі шағын, теңгерімделген мәтіндер корпусы.

Аударма – қандай да бір тілдегі мәтіннің басқа бір немесе бірнеше тілге аударылған нұсқасы.

Ауызша тіл корпусы – кино, радио т.б. көздерден таспаға жазылып алынған ауызша туындаған мәтіндер корпусы.

Ашық дауысты дыбыс – тілдің төмендеп барып көтерілуі арқылы жасалатын дауыстылар, оларды айтқанда жақ кең ашылып, иек төмендейді. Қазақ тіліндегі ашық дауыстылар: а, ә, о, ө.

Белгіленім – мәтінге және оның компоненттеріне (бөліктеріне) арнайы тәгтер (белгіленімдер) қою: әдетте мәтін элементтерін лексикалық, грамматикалық және т.б. ерекшеліктері бойынша сипаттайтын лингвистикалық белгіленімдер; сондай-ақ экстралингвистикалық белгіленімдер (мәтін және автор туралы ақпарат: автор, атауы, жылы, шыққан жері, мәтіннің жанры, тақырыбы т.б.).

Белгіленім жүргізу – мәтін бірліктеріне алдын ала шартты түрде арнайы қабылданған тәгтерді «тіркеп» жазу әрекеті.

Билік атаулары – ел басқару, әкімшілік басқару атақ-лауазымдарын білдіретін сөздер.

Билік институты – мемлекеттің саяси және әлеуметтік жағдайын тек құқықпен ғана емес, рухани-мәдени салт-дәстүр негізінде де ұштастыра әрі

теңестіре отырып қамтамасыз ететін рөлі ерекше, мемлекеттік басқару жүйесімен тығыз астасып жатқан бірден-бір құрылым.

Вариант <лат. *varians* – өзгеретін, ерекшеленетін> – айтылым кезінде сөздердің дыбыстық өзгеріс нәтижесінде сол тілдің дыбыс қорында бар басқа бір дыбысқа ұқсап айтылуы.

Вариант дыбыс/өзгелік реңк – айтылу кезінде тілдегі басқа дыбысқа ұқсап кететін, дыбыстың уақытша түрленімі.

Вариация/бөгде реңк <лат. *variatio* – өзгеріс, ауысу> – айтылым кезінде сөздердің дыбыстық өзгеріс нәтижесінде сол тілдің дыбыс қорында жоқ, басқа бір дыбысқа ұқсап айтылған дыбыс түрленімі. Гидроним – топонимнің бір түрі, су объектілерінің жеке атауы. Мысалы, Жайық, Талас, Шалқар, Ертіс, Балқаш т.б.

Ғалам бейнесі – адам және оның болмысының, адамның әлеммен қарым-қатынасының, адамның өмір сүру жағдайының өзіндік ерекшеліктерін білдіретін іргелі де күрделі ұғым.

Дамма – араб жазулы мәтіндерінде қазақ тіліндегі «ұ», «ү», «о», «ө» еріндік дауысты дыбыстардың мәнін білдіретін, әріпүсті үтір арқылы таңбаланатын диакритикалық белгі.

Диалектілік мәтіндер корпусы – жергілікті аумақтағы сөйлесім дискурсы жазылып алынған мәтіндерді грамматикалық тұрғыдан сипаттайтын, сонымен қатар сөздердің жергілікті мағынасы мен морфологиясын сипаттайтын корпус.

Дыбыс қоры – бір фонеманың өздік, өзгелік, бөгде реңктерінің жиынтығы.

Езулік дауыстылар – айтылуда еріннің кейін тартылып, езудің жиырылуы арқылы жасалатын дауысты дыбыстар: а, ә, е, ы, і.

Еріндік дыбыс – артикуляциясы бойынша еріннің дөңгеленіп алға қарай сүйірленуі немесе бір-біріне жабыса тиюі арқылы жасалатын дыбыстар.

Ерін үндестігі – сөйлеуде алғашқы буындағы еріндік дауыстының әсерімен кейінгі буындардағы әлсіз (қысаң) дауыстылардың еріндікке айналуы. Мысалы, бұрынғы – бұрұңғұ, өрік – өрүк, күрек – күрөк, көлеңке – көлөңкө және т.б.

Жадид жазуы – дауысты дыбыстар харакаттар (фатха, кесра, дамма) арқылы емес, арнайы таңбамен берілетін араб жазуының бір түрі.

Жарнама – бір нәрсенің (тауар, қызмет түрлері, т.б.) сұранысын арттыру, тұтынушылар назарын аудару мақсатында жарияланатын хабарлама, ақпарат.

Жартылай балама – аударылатын тілде түпнұсқа тілдің бірегей баламасы.

Жіңішке езулік дауысты – тілдің жоғары қалпымен ерін қатысуынсыз езу емеурін арқылы жасалатын сөз құраушы дауысты дыбыс.

Жіңішке еріндік дауысты – тілдің жоғары қалпымен, ерін қатысуымен жасалатын сөз құраушы дауысты дыбыстар.

Зооним – жануарлардың атауын білдіретін онимнің бір түрі.

Кесра – араб жазулы мәтіндерінде қазақ тіліндегі «ы», «і» жіңішке, қысаң дауысты дыбыстардың мәнін білдіретін, әріпасты көлбеу сызықша арқылы таңбаланатын диакритикалық белгі.

Коммуникация – байланыс, қарым-қатынас тәсілдері, ақпараттарды беру және қабылдау жолдары.

Конкорданс – қажетті сөздің дереккөзге сілтеме жасалған барлық контекстегі қолданысының тізімі.

Корпус-менеджер <ағыл. corpus manager> – корпустан қажетті ақпаратты, статистикалық мәліметтерді тауып алуға арналған компьютерлік бағдарлама бойынша жұмыс істейтін арнайы іздеу жүйесі.

Корпустық лингвистика – компьютерлік технология негізінде лингвистикалық корпустарды (мәтіндер корпусын) құрастыру мен оны пайдаланудың жалпы ұстанымдарын зерттейтін компьютерлік лингвистиканың бір саласы.

Көркем аударма – қандай да бір тілдегі әдеби шығарманың басқа бір немесе бірнеше тілге аударылған нұсқасы, көркем әдеби шығармашылықтың бір саласы.

Қадим жазуы – дауысты дыбыстар таңба арқылы емес, диакритикалық белгілермен, яғни белгілі бір харакаттар (фатха, кесра, дамма) арқылы берілетін араб жазуының бір түрі.

Қатыстық балама – түпнұсқа тілі сөзінің жалпы мағынасын беретін, бірақ оның стильдік жағынан мағынасын толық ашпайтын балама.

Лексикалық тіркесімділік – сөздің өзіндік мағынасының ұғымдық және заттық, сезімдік тұрғыдан үйлесе алу не тіркесім құрай алу мүмкіндігі.

Лексикалық мағына – зат, құбылыс, іс-әрекеттің санада бейнеленіп тұрақталған ұғымының мазмұны.

Лексикалық минимум – тіл меңгерудің әрбір деңгейіне сәйкес келетін (қарапайым, базалық, орта, ортадан жоғары және жоғары деңгейлер) минималды сөздер мен сөз тіркестерінің жиыны.

Лексика-семантикалық белгіленім – сөздерді мағыналық топтарға ажырататын және оларға тән лексика-грамматикалық категорияларды қамтитын лингвистикалық белгіленім түрі.

Лексика-семантикалық субтоп – сөздердің ажыратушы семасы негізінде терең мағыналық бөлінісін көрсететін кіші лексика-семантикалық топ.

Лексика-семантикалық ірі топ – сөздерді ортақ семасы негізінде біріктіретін ірі мағыналық немесе тақырыптық топ.

Лемма – сөздің бастапқы тұлғасы.

Лемматизациялау – түрленген сөзформалардан сөздің алғашқы формасын (яғни реестр сөзді) бөліп алу процесі.

Лингвистикалық белгіленім – мәтіндегі әрбір лексикалық бірлікке орфографиялық, фонетикалық, лексикалық, грамматикалық тұрғыдан берілетін лингвистикалық ақпарат.

Лингвистикалық (тілдік) мәтіндер корпусы – құрылымдалған, белгіленімдер қойылған, филологиялық тұрғыдан ақпараттандырылған, нақты лингвистикалық мәселелерді шешуге бағытталған, автоматты түрде жұмыс істейтін, үлкен көлемдегі тілдік деректер жиынтығы.

Макротопоним – ірі физикалық-географиялық объектілердің (тау жоталары, үлкен ойпаттар, мұхиттар, т.б.) атаулары. Мысалы: Сарыарқа, Ақжайық, Нарын, Бетпақдала, Қызылқұм.

Мәдени коммуникация – әр мәдениеттің өзіндік ерекшелігін құрметтеу негізінде мәдениеттің барлық саласында жүзеге асырылатын қарым-қатынас құралдары, ақпаратты бір мәдениеттен екінші мәдениетке тасымалдау.

Мәдени коннотация – белгілі бір заттық атауыштың бойындағы қасиет пен ерекше белгісін әлемдік немесе ұлттық мәдени болмыспен теңестіре, сәйкестендіре сипаттау арқылы пайда болатын үстемелі мағыналар көрінісі.

Мәдени сабақтастық – ұрпақтан-ұрпаққа жетістіктер мен құндылықтарды, тәжірибе мен дағдыны жалғастырып отыратын қоғам дамуының барлық сатысында кездесетін заңды құбылыс.

Мәдени-семантикалық белгіленім – сөздердің мағыналық құрылымында кодталған ұлт мәдениеті туралы білімдердің жүйесі.

Мәтіндер корпусы – шектелген көлемдегі мәтіндер жиыны.

Метабелгіленім – мәтіндердің атауы, авторы, жылы, стилі, жанры, формасы, аудиториясы, шыққан жылы, көлемі, сөзқолданыс саны т.б. бойынша сипаттайтын компьютерлік бағдарлама.

Метамәтіндік белгіленім – мәтін туралы берілетін сыртқы ақпаратты қамтитын (мәтін авторы, атауы, жылы, стилі, жанры т.б.) экстралингвистикалық белгіленім.

Метасипаттама – мәтінге қатысты мәліметтерді (авторы, тақырыбы, жариялануы, баспасы, жанры т.б.) сипаттайтын белгіленім.

Микротопоним – кең қауымға беймәлім географиялық нысандардың (көл, бұлақ, құдық, қоныс, т.б.) атаулары.

Мифологиялық шығарма – аңыз, ертегілерге негізделген шығарма түрі.

Мифоним – мифтердегі, эпопея, ертегілердегі, жырлардағы кез келген ономастикалық кеңістік аясындағы атау, аңыз-әңгімелерде, мифтерде сақталған тылсым, қасиетті саналатын, адамдар оларды шынайы бар деп түсінетін атаулар, мыс., Құлагер, Шалқұйрық, Қобыланды батыр, Қамбар ата, Шопан ата, Зеңгі баба.

Мифтік таным – мифтік сана дәуірінде дүниенің миф арқылы біртұтас, бөлшектенбеген, синкретті, әмбебап қалпында бейнеленген қоғамдық сананың түрі.

Морфологиялық белгіленім – сөз таптарын және оларға тән грамматикалық категорияларды қамтитын лингвистикалық анализатор (талдауыш); мәтін бірліктерін морфологиялық ерекшеліктеріне қарай сипаттайтын белгіленімдер (лемма, сөз табына қатысы, грамматикалық категорияларға қатысы).

Ойконим – топоним түрі; елді мекен (ауыл, село, кент, қала) атауы.

Оқу корпусы – лингвистикалық белгіленімдері мектеп бағдарламасына лайықталып жасалған, әдетте омонимдер ажыратылған оқу-әдістемелік корпус; тіл үйренушілердің деңгейіне сай оқытатын мәтіндердің электронды қоры.

Ономастика <грек *onomastike* – ат беру өнері> – тіл білімінің жалқы есімдерді зерттейтін саласы. Жалқы есімдерге кісі, жануарлар аттары, халық, ұлт, ру-тайпа атаулары, жер-су, аспан денелері жатады.

Ономастикалық кеңістік – белгілі бір жалқы есімдердің таралу аймағы немесе жалқы есімдердің тобы; ономастикалық кеңістік топонимия, антропонимия, зоонимия, теонимия сияқты топтардан тұрады.

Ономастикалық ішкорпус – электронды түрдегі ономастикалық мәліметтер қамтылған ақпараттық-анықтамалық жүйе.

Ороним – жер үсті қабатындағы нысандардың (тау, қырат, шың) атаулары. Мысалы: Алатау, Хан Тәңірі, Алтай, Орал, Қаратау.

Орфоэпия – тілдің айтылым нұсқасын және оның тілдік дәстүрге сай дұрыс айтылымын іріктеп көрсететін фонетика саласы; тілдегі айтылым нормасының жүйесі.

Параллель корпус – екі немесе одан көп тілдегі сөздер мен сөз тіркестерінің аудармасын тауып алуға болатын екі немесе одан көп тіл бойынша теңестірілген мәтіндер базасы негізінде әзірленген корпус.

Параллель мәтін – қандай да бір тілде жазылған түпнұсқа мәтін және оның басқа бір немесе бірнеше тілге аударылған мәтіні.

Параллель мәтіндерді теңестіру – қандай да бір тілде жазылған түпнұсқа мәтіндегі сөзді немесе сөйлемді аударма тілдің мәтініндегі сөзбен немесе сөйлеммен автоматты түрде немесе қолмен сәйкестендіру.

Парсинг – лексемалардың (сөздердің) тізбекті байланысын формальді грамматикамен салыстыру процесі. Нәтижесі синтаксистік қатынас (тәуелділік-дерезависимость) болып табылады.

Поэтикалық мәтіндер корпусы – лексикалық және грамматикалық қана емес, сонымен қатар өлең стиліне қатысты ерекшеліктерді де сипаттайтын және өзіндік ерекшеліктеріне қарай іздеуге болатын корпус.

Просодикалық белгіленім – корпусқа енгізілген мәтін бірліктерінің екпін мен интонацияға қатысты ерекшелігін сипаттайтын белгі (тәг).

Просодикалық бірліктер <просодика – грек. *προσῳδία* (*prosodia*) – екпін> – сөз ағымының тізбекті бірліктерінің (буын, сөз тіркесі, синтагма, фраза, мәтін) үстінен түсетін барлық бөлігінде көрінетін, интонация арқылы ажыратылатын жалпы категориялық (тиянақты/тиянақсыз, хабарлау/сұрау, т.б.) мағыналарды айқындауға қатысатын суперсегментті фонетикалық бірліктер (сөйлеу әуені, екпін, тембр (әуез), ырғақ, т.б.).

Репрезентаттық принцип – мәтіндерді корпусқа енгізудегі бір тілдің барлық қабатын (стильдер, жанрлар, тақырыптар) қамтитын таңдама мәтіндерді жинақтау ұстанымы.

Семантика <көне грекше: *σημαντικός* – танбалаушы, білдіруші> – тіл және тіл бірліктері (сөз, грамматикалық тұлға, сөз тіркесі, сөйлем) арқылы білдірілетін хабарды, заттар мен құбылыстардың мән-мазмұнын зерттейтін тіл білімінің саласы, семиотиканың негізгі бөлімдерінің бірі.

Семантикалық белгіленім – корпусқа енгізілген мәтіндегі сөздердің мағынасына, категориясына, тақырыптық топқа, т.б. сипаттарына қарай қойылатын белгіленім.

Семантикалық тәгтер – мәтіндегі сөздің немесе сөз тіркесінің қандай да бір семантикалық категорияға немесе оның шағын ішкі топтарына қатысын сипаттайтын белгі.

Синтаксистік белгіленім – корпусқа енгізілген мәтіндегі лексикалық бірліктер мен синтаксистік құрылымдардың қатынасын, байланысын сипаттайтын белгіленімдер (мысалы, етістікті тіркес, есімдікті тіркес т.б.).

Слоган – тұтынушы назарын аударуға бағытталған жарнамадағы ұран сөздер.

Сөздік қор – сөздік құрамның ең тұрақты бөлігі, ондағы сөздердің негізгі ұйтқысы.

Сөзжасамдық белгіленім – сөздің жасалу жолына, тәсіліне қатысты лингвистикалық ақпарат беретін белгіленім түрі.

Сөйлеу (дыбыстық) корпусы – дыбыстық файлдар мен жазып алынған мәтіндер транскрипциясының дерекқоры түріндегі мәтіндер корпусының бір түрі. Сөйлеу технологиясында сөйлеу корпустары акустикалық модельдер жасау үшін қолданылады, оларды табиғи тілдің ауызша формасын тану механизмдерінде қолдануға болады. Тіл білімінде сөйлеу корпустары фонетика, диалектология, конверсиялық талдау және басқа салаларды зерттеу үшін қолданылады.

Стемминг – сөздің негізін контекссіз табу.

Тарихи корпусстық лингвистика – компьютерлік технологияларды қолдана отырып, тарихи мәтіндер корпусын жасау мен оны пайдаланудың жалпы принциптерін зерттейтін қолданбалы тіл білімінің жаңа бір саласы; қазақ тілінің ескі кезеңдерден бері жарық көрген әртүрлі графикалық жүйедегі тарихи ескерткіштер мәтіндерінің лингвистикалық сипаттамасы бар электронды жиынтық; сөздердің даму тарихы туралы кең ауқымды ақпарат алуға болатын платформа.

Тег (белгіленім тілі) – гипермәтіннің белгіленім тілінің элементі (мысалы, XML, HTML); аудио және видеофайлдардағы дыбысжазбаны авторы, аты бойынша қалай табу жолдарын сипаттайтын ақпараттық белгі (мысалы, ID3-тег).

Теоним – құдай, әулие атаулары және шығармалардағы діни атаулар.

Терең аннотацияланған корпус – мәтіндегі сөйлемдердің морфологиялық және синтаксистік белгіленімдері берілген корпус.

Толық балама – бір мағыналы сөздің мағынасын толық жеткізетін балама.

Топоним – онимдердің бір түрі; топонимдерді, көбінесе, жер-су атаулары деп те атаймыз; топонимдер – жер бетіндегі табиғи объектілердің және адамдардың табиғатты игеру үрдісі нәтижесінде қалыптасқан объектілердің атаулары; топонимдер өз ішінен ойконимдер, гидронимдер, оронимдер, урбанонимдер, некронимдер т.б. түрлерге бөлінеді.

Төте жазу – қазақ әліпбиінің реформаторы Ахмет Байтұрсынұлының араб әліпбиі негізінде жетілдірген ұлттық жазу.

Транскрипция <лат. trans – арқылы; seribo – сызамын, жазамын> – белгілі бір тілдегі дыбыстарды дәл бейнелеп көрсету үшін пайдаланатын шартты таңбалар жүйесі.

Тілдің автоматтануы – мәтінді не сөзді оңай және тікелей тануды қамтамасыз ететін электронды ресурстардағы тілтанымдық ақпараттардың цифрландырылуы.

Тіркесімділік – тілдік бірліктердің арасындағы жоғары деңгейдегі бірліктер құрудағы тіркесімінің синтагматикалық қатынастарын сипаттайтын қасиеті.

Урбаноним – қалаішілік нысан атаулары. Мысалы: агороним – алаң атаулары, годоним – көше атаулары, хороним, эмпороним – дүкен атаулары.

Ұлттық баспасөз – белгілі бір ұлтқа тән, сол ұлттың мәдени-рухани және басқа да барлық саласына қатысты өзекті мәселелерді көтеретін бұқаралық ақпарат құралы.

Үндестік заңы – сөздің бастан-аяқ біркелкі әуезбен айтылуын, сөздің өн бойында әуездің сақталуын реттейтін фонетикалық заңдылық.

Факсимиле <лат. fac simile – ұқсау> – көне немесе орта дәуірдегі қолжазбалардың немесе ескі кітаптардың бастапқы қалпына ұқсас етіп басылған көшірмесі.

Фатха – араб жазулы мәтіндерінде қазақ тіліндегі «а», «ә» жіңішке, езулік дауысты дыбыстардың мәнін білдіретін, әріпүсті көлбеу сызықша арқылы таңбаланатын диакритикалық белгі.

Фитоним – өсімдік атауларын білдіретін онимнің түрі. Мысалы, раушангүл, лилия.

Хороним – территория, аймақ, облыс, аудан атауы. Мысалы: әкімшілдік, қалалық, табиғи хоронимдер: Шығыс Қазақстан, Орта Азия, Еуразия, Алматы облысы, Шымкент қаласы.

Іздеу жүйесі – қолданушының сұрауын қамтамасыз ететін іздеу бөлігі.

Экстралингвистикалық белгіленім – корпусқа енгізілетін мәтіннің сыртқы сипатынан ақпарат беретін белгілер (библиографиялық, әлеуметтік, тақырыптық, типологиялық сипаттама, құрылымдық сипаттама (мәтін, тарау, бөлім, абзац, сөйлем т.б.), мерзімі, орындаушылар, дереккөзі т.б.).

Этнолингвистикалық белгіленім – халықтың этногенезіне, тұрмыс-салты мен әдеп-ғұрпына, мекенжайына, басқа халықтармен тарихи-мәдени байланысына, күнделікті тұрмысына, материалдық және рухани мәдениетіне қатысты қолданылатын сөздерге берілетін лингвистикалық ақпарат.

ҚЫСҚАША АҢДАТПА

Монография Қазақ тілінің ішкорпустарын әзірлеу тәжірибесін сипаттауға арналған 11 тараушадан тұрады.

1. Қазақ тілінің ұлттық корпусын жасаудың концептуалды негіздері. Бұл тараушада Қазақ тілінің ұлттық корпусын жасауға арналған концепция мазмұны мен А.Байтұрсынұлы атындағы Тіл білімі институтындағы әзірлеу кезеңдері сипатталады.

2. Ауызша ішкорпус. Сөйлеу тілі – сол ұлттың сөйлеу дәстүріне табан тірейтін, сөйлеу дәстүрін танытатын тіл болғандықтан, бұл мәтіндер ұлттық корпусының мазмұнын байытатын мәтіндер болып табылады. Ауызша тілдің іш корпусына күнделікті тұрмыстық қатынастағы ауызекі сөйлеу стилі (үй-іш, отбасы, туған-туыс, көршілер арасындағы диалогтар, дүкен, көлік, қоғамдық орындардағы тілдесулер, бейресми сипаттағы телефон арқылы сөйлесулер), бейресми әңгіме, сұхбаттардағы машықсыз, дәстүрлі сөйлеулер, ауызша шаршысөз, дәріс, конференция, дөңгелек үстел, кеңес, семинар, сұхбаттардағы алдын ала жоспарланған тақырыпқа сөйленген сөздер, кино, сахналық қойылымдағы дискурстар, ауызша сөйленген машықты, машықсыз діни дискурстар, телерадиодағы алдын ала тақырыбы белгілі бағдарламалардағы машықсыз сөздер алыну қажеттігі дәлелденеді. Себебі машықсыз сөйлеу тілі шынайы, табиғи тіл болып табылады. Осыған байланысты мақалада ауызекі сөйлеу тілінің ұлттық ерекшелігі сипатталады. Бұл ретте бірыңғай қазақ ұлты қоныстанған өңірлердегі жасы үлкен қариялардың дәстүрлі сөйлеу тілі талданады.

Қазақ тілінің ауызша ішкорпусы: мәтіндер корпусындағы фонетикалық, просодикалық белгіленімдер әзірлемесі. Зерттеуде мәтіндерді автоматтандыру үшін қазақ тіл білімінің сегментті және суперсегментті фонетикаға қатысты білімдер жинақталып, жүйеленіп фонетикалық және просодикалық белгіленімдер әзірленді. Қазақ тіліне тән әрбір дыбысқа фонетикалық, просодикалық сипаттама беріліп, интонация компоненттеріне шартты таңбалар қойылды. Мәтінге фонетикалық және просодикалық белгілерді қоюдың жолдары қарастырылды.

Ауызша ішкорпус базасындағы мәтіндерге просодикалық белгіленім қойылып, сөйлеушінің өз сөзі мен оның сөзінің орфоэпиялық нормасы жасалды. Сондай-ақ ауызша корпус базасы сөздің фонетикалық-просодикалық ерекшелігін көрсетуде көптеген зерттеуге негіз болады.

Ауызша ішкорпус базасындағы мәтіннің фонетика-фонологиялық және интонациялық транскрипциясы. Зерттеуде фонетика-фонологиялық транскрипцияның қазіргі тіл біліміндегі рөлі мен оның қолданыс аясы сөз болады. Қазіргі тіл білімінде мәтіндерді транскрипциялау барысында просодикалық жағы ескерілмей келеді. Зерттеу барысында ең алғаш қазақ тіл біліміндегі интонация мәселесін жан-жақты зерттеген академик З.М. Базарбаеваның зерттеу еңбектері басшылыққа алынды. Сондай-ақ мәтін талдау барысында, негізгі 8 интонома модельдері берілді. Яғни фонетикалық транскрипциямен қатар просодикалық ерекшеліктерді анықтау мақсатында мәтіннің интонациялық транскрипциясы да берілетін болады. Сондай-ақ мәтінді транскрипциялау бары-

сында әрбір сөйлем синтагмаға, ырғақтық топқа бөлініп, дауыс үдемелігі, әуен, кідіріс, қарқын сияқты компоненттері арнайы белгілермен беріліп, сипаттама жасалады. Әрбір сөйлемнің интонемаларының моделі беріледі.

Орфоэпиялық конвертер жасау тәжірибесі. Конвертер – бір типтегі файлдарды басқа түрдегі файлдарға түрлендіретін арнайы бағдарлама. Орфоэпиялық конвертер – орфографиямен жазылған сөзді, сөз тіркесті, сөйлем мен мәтінді орфоэпиялық нормаға сай жазып беретін бағдарлама.

Институттың қолданбалы лингвистика және фонетика бөлімі орфоэпиялық конвертер жасау ісіне жауапты болды. Осыған дейінгі фонетикалық еңбектерді зерттей келе, толықтырылған орфоэпиялық ережелердің алгоритмін әзірледі. Жалпы саны 183 алгоритімді нақты мысалдармен дәлелдей отырып, бекітті.

Жасалған орфоэпиялық конвертер бірнеше маңызды мәселені шешуге қауқарлы. Біріншіден, қазақ тілінің ұлттық корпусының (<https://qazcorpus.kz/>) ішкі белгіленімін толықтырып, сөздерді буынға дұрыс бөледі. Екіншіден, ауызша корпус базасына енетін мәтіндерді орфоэпиялық нормаға сай жазып береді. Үшіншіден, жеке өнім ретінде қолданысқа ұсынылады. Сонымен қатар сөйлеу технологияларын жасауда автоматтанған орфоэпиялық нормалардың, сөздіктердің болуы сапалы нәтиже үшін қажет.

3. Ономастикалық ішкорпус. Ахмет Байтұрсынұлы атындағы Тіл білімі институты Ұлттық корпусдың дербес бір бөлігі саналатын ономастикалық ішкорпусты алғашында топонимика саласы бойынша құрастыруды бастады. Бұл тақырыпта ономастикалық ішкорпусты құрастырудың бағыты, ұстанымдары талқыланады. Сол арқылы ономастикалық ішкорпусты құрастырудың негізгі критерийлері айқындалады. Қазақ тілінің ұлттық корпусы мен әлемдегі ірі ірі корпусардың тәжірибесі, ерекшеліктері сипатталады. Топонимдік белгіленімдер әзірлеуге қатысты салыстырулар жасалып, ұсыныстар айтылады. Шетелдік корпусар тәжірибесінен байқағанымыздай, неғұрлым ішкорпустар көп болса, соғұрлым ұлттық корпус кемелдене түседі. Ономастикалық ішкорпустың пайда болуы ономаст ғалымдардың зерттеулеріне көптеген жаңа мүмкіндіктер алып келеді. Өйткені корпусан жер-су аттары кездесетін мәтіндерді жылдам табуға мүмкіндік туады. Бұл тек іздеу жұмысын жеңілдетіп қана қоймай, ономастика саласындағы кешенді зерттеулерге де өз септігін тигізеді. Ономастикалық ішкорпус географиялық атауларды тек бекіту және стандарттау тұрғысынан ғана емес, сонымен бірге топонимдерді жүйелеу тұрғысынан, оларды сақтау және ғылыми зерттеу мақсатында осы ұлттық қорға қол жеткізу және танымдық қызығушылықты қанағаттандыру тұрғысынан да сөзсіз қажет.

Мифонимдер мен теонимдерді қазақ тілінің ұлттық корпусына енгізу технологиясы. Қазақ тіл білімінің ономастика саласы бойынша біраз зерттеулер жүргізіліп, бірқатар диссертациялық жұмыстар қорғалды, атаулар сөздіктері құрастырылды. Алайда, кейбір мифонимдер мен теонимдерге, әсіресе мифтерге негізделген, атаулар бойында сақталып қалған ежелгі дәуірлік этностардың қоршаған әлем бейнесін қабылдаудан туындаған дүниетанымдық көзқарастарының сырын ашу бағытындағы зерттеулерде бірізділік жоқ. Соған орай, бұл тақырыпта негіз болған мифтік және діни атауларды көпшілікке

қолжетімді интернет желісінде жан-жақты лингвистикалық талданған белгіленімдер арқылы корпусқа енгізу технологияларын жинаған материалдарға қатысты салыстыра отырып, осы тектес мифонимдер мен теонимдер атауларына қатысты ортақ пікірлер мен тұжырымдар талданады. Мифонимдер мен теонимдерді қазақ тілінің ұлттық корпусына енгізу технологиясы туралы ақпарат беріледі. Бұл қазақ тілінің мифтік, діни кеңістіктерін рухани құндылықтар жүйесі арқылы жаңғыртуға, тілді оқыту үдерісінің тиімділігін арттыруға ықпал етеді.

4. Тарихи ішкорпус әзірлеудің әлемдік тәжірибесінен. Ұлттық корпус ұғымы тек синхрониялық емес, сонымен қатар диахроникалық зерттеулердің құралы болып табылады. Осыған орай аталмыш тараушада қазақ тілінің ұлттық корпусының аясында «тарихи ішкорпусты» құрастыру қажеттілігі айтылады. Себебі «тарихи ішкорпус» – V-XX ғғ. қамтитын жазба мұралардың тілін, тарихын, мәдениетін, әдебиетін зерттеуде кез келген пайдаланушыға бірден-бір пайдалы, аса қажетті құрал болып есептеледі. Қолданбалы лингвистика саласында «тарихи ішкорпус» тұңғыш рет әзірленіп жатыр. Алайда түрлі графикамен жеткен қолжазбаларды цифландырудың өзіндік қиындықтары бар. Зерттеу жұмысында тарихи ішкорпусты әзірлеудің әлемдік тәжірибедегі құрылымы мен практикалық қолданысы зерттеледі. Өзге елдердің тәжірибелерін ескере отырып, қазақ тілінің тарихи корпусын әзірлеудің негізгі атқарылатын жұмыстары нақтыланады. Әр дәуірдегі жазба ескерткіштер мәтіндер базасын анықтау, жинау; жинақталған материалдардың сапасы мен құрамын сұрыптау, топтастыру, өңдеу; оларды корпусқа енгізу және демонстрациялау мәселелері, әрбір жазба мұраға метабелгіленімдердің құрылымы айқындалады. Ұсынылған зерттеу жұмысын кез келген тілдің тарихи ішкорпусын әзірлеуде пайдалануға болады.

Орыс тілінің ұлттық тарихи корпусының тәжірибесі мен түркілік лексикалық қабаттың берілуі. Тіл мен жасанды интеллект, тіл мен компьютерлік қолданбалы лингвистика саласындағы байланыс қазіргі таңда қарқын алып дамып келе жатыр. Осы байланыстың жалғасы әрі жаңа белесі ретінде тілдегі ескі жазба мұраларды компьютерлік қолданбалы лингвистиканың объектісі ретінде қарастыру бүгінгі заманның талабы. Дамыған әлем лингвистикасында қолданбалы лингвистика, корпус, соның ішінде тарихи корпус жасау тәжірибесі қалыптасып, өз жемісін, көптеген ғылыми зерттеулерге жаңаша қарқын, зор мүмкіндіктер беріп жатыр. Осы мақсатта біз қолданбалы лингвистикасы мен корпус жасауы жақсы қалыптасып дамыған, орыс тілінің тарихи корпусын зерттеп, ондағы ескі тіл бірліктерінің берілу тәжірибесі мен түркілік лексикалық қабаттың көрініс табуын қарастырдық. Сонымен қатар әлемдік тарихи корпус жасау тәжірибесін бағамдау үшін Армян тілінің ұлттық корпусы да қарастырылды. Қазақ тілінің тарихи корпусын жасау, жетілдіру барысында тарихи сөздік, тарихи ономастика, тарихи терминология, тарихи орфография, тарихи грамматика, тарихи сөзжасам, тарихи фразеология, тарихи тіркесімділік, тарихи теология, ескі діни мәтіндер мен тарихи жиілік корпустарын жасауға болатындығын ұсынамыз. Тарихи корпусты жасау – кез келген тілдің жаһандануға жұтылып кетпеуінің бір тетігі. Тарихи корпус – кітапхана мен оқырманның арасындағы замануи цифрлық алтын көпір.

Тарихи ішкорпус мәтіндерінің лексика-грамматикалық белгіленімі.

Лингвистикалық белгіленім – мәтіндегі әрбір лексикалық бірлікке орфографиялық, фонетикалық, лексикалық, грамматикалық тұрғыдан берілетін лингвистикалық ақпарат. Алайда тарихи ішкорпус мәтіндеріне лингвистикалық белгіленім әзірлеу – терең ізденісті қажет ететін аса күрделі жұмыстардың бірі. Себебі тарихи ішкорпусқа енген мәтіндер базасының басым көпшілігі араб графикасымен жазылған. Ал араб графикасымен жазылған мәтіндер кирил графикасына әртүрлі транскрипцияланған. Тарихи ішкорпусты әзірлеуде басқа ішкорпустарға қарағанда теориялық және техникалық тұрғыдан өзіндік қиындығы бар. Осығай орай тарихи ішкорпус мәтіндеріне лингвистикалық белгіленімдерді қою мәселесі назарға алынады. Бұл ретте орыс тілінің тарихи ішкорпусын әзірлеу тәжірибесі қарастырылады; әр кезеңде жазылған араб-графикалы мәтіндерінің транскрипциясы талданады; транскрипцияланған сөздердің варианттары анықталады; транскрипцияланған мәтіндер үшін лексика-грамматикалық белгіленімдер айқындалады; транскрипцияланған мәтіндерге арналған лексика-грамматикалық іздеу жүйесінің механизмдері сипатталады.

5. Параллель ішкорпус. Аударматану ғылымының басты ұғым-түсініктері. Аударма ғылымы, аударматану, аударма қызметі деп аталатын маңызды да өзекті, әрі қызықты саланың тамыры қай халықтың болмасын тарихының тереңінен бастау алатындығы белгілі. Халық болып қалыптасқан әрбір ұлттың мәдениетінде, әсіресе жазу-сызуы қалыптасқан кезеңдерден бастап аударма ісі қоғамдық өмірдің ажырамас бөлшегіне айналды; ол туралы көзқарас-пікірлер айтыла бастады. Бұндай көзқарастарда көбіне-көп қоғамдағы, мәдени өмірдегі аударманың орны қандай, мәдениетке қосар үлесі қандайлық, ерекшеліктері неде деген сауалдарға жауап іздестірілді. Ал аударманың лингвистикалық теориясы деген ұғымның қалыптасуы 20-сыншы ғасырдың жиырмасыншы жылдарына тиесілі; алғашқы кезде бұл термин көркем әдебиеттің аудармасы деген ұғыммен тығыз байланыста қарастырылды.

«Қазақ және орыс тілдерінің параллель ішкорпусы: әзірлемесі, қолданылуы және мәселелері» атты тақырыпта параллель ішкорпустардың түрлеріне қысқаша сипаттама берілді. Қазақ тілінің ұлттық корпусының құрамында әзірленіп жатқан параллель ішкорпуста М.Әуезовтің «Абай жолы» роман-эпопеясының қазақ тіліндегі түпнұсқа мәтіні мен оның А.Ким жасаған орыс тіліндегі мәтінінің аудармасы абзац (сөйлем) деңгейінде қолмен теңестірілді.

Параллель ішкорпусты әзірлеу барысында Microsoft Office Excel, Notepad++, Python, Django, MySQL бағдарламалық құралдары қолданылды.

Параллель ішкорпус теңестірілген мәтіндер базасынан және іздеу жүйесінен, енгізілген мәтін туралы ақпарат (метабелгіленім) мынадай параметрлерден тұрады: *автор, аудармашы, шығарма атауы, аударма атауы, шығарманың басылған мерзімі, аударылған мерзімі, түпнұсқа тіл, аударма тіл*. Іздеу жүйесі ізделетін сөзді сөз, сөз тіркес, сөйлем, бас әріп (қазақ және орыс тілдеріндегі) параметрлері арқылы табуға мүмкіндік береді.

«Абай жолы» романындағы бейвербалды амалдардың аударылу ерекшеліктері. «Абай жолы» роман-эпопеясы – қазақ халқының тұрмысы

мен тарихынан мол дерек беретін мәдени-тілдік құндылығы жоғары көркем шығарма. Роман XX ғасырдың екінші жартысынан бастап бүгінгі күнге дейін әлемнің көптеген тілдеріне үздіксіз аударылуда. Алайда шығарманың аудармалары түпнұсқа, яғни қазақ тілі арқылы емес, дәнекер тіл – орыс тілі арқылы жасалып жатыр. Екіншілік аударма барысында түпнұсқадағы автор идеясы ғана емес, мәтіннің көркемдік қуаты және тілдік, танымдық ерекшеліктері сақталуы – даулы мәселе. Сол себепті «Абай жолы» романындағы тілдік бірліктердің аудармасын лингвистикалық тұрғыдан талдау әрдайым өзекті. Халықтың ұлттық ерекшелігін танытатын бейвербалды амалдар да кез келген көркем мәтінде әрі саны жағынан, әрі сапасы, яғни экспрессивтік, ақпараттық мәні жағынан маңызды орынға ие. Зерттеуімізде түпнұсқа мәтіндегі параэлементтер аударма мәтіндегі баламаларымен салыстырыла отырып талданды. ЫМ-ишараттардың *аударылмаған тілдік бірліктер (лакуналар), сөзбе-сөз аударылған тілдік бірліктер, сапалы аударылған тілдік бірліктер, теңестірілген тілдік бірліктер, еркін түрде аударылған тілдік бірліктер, жартылай аударылған тілдік бірліктер* сынды алты түрлі жолмен аударылғаны атап өтіліп, мысалдармен дәйектелді.

Параллель мәтіндер корпусының лексикографиялық әлеуеті тақырыбы параллель мәтіндер корпусының лексикографиялық әлеуетін зерттеуге арналған. Жұмыста талдау әдістемесі мен құралдарына, параллель мәтіндерден лексикографиялық тұрғыдан түпнұсқа ақпаратты бөліп көрсетуге баса назар аударылады. Нәтижелер лексикография туралы түсінігімізді байытады және параллель корпус пен табиғи тілді компьютерлік өндеудегі қазіргі зерттеулердің болашағын ашады.

6. Оқу корпусы. Қазақ тілінің оқу корпусын әзірлеу мәселелері. Бұл тақырыпта оқу корпустарын әзірлеудің әлемдік тәжірибесіне шолу жасалады. Ағылшын, орыс тілінің Оқыту корпустарының мәтіндік базасында қандай мәтіндер қамтылған, тілүйренушілер үшін қандай лингвистикалық ақпараттар берілген т.б. мәселелері зерттеледі. Сонымен қатар А.Байтұрсынұлы атындағы Тіл білімі институтында әзірленіп жатқан Қазақ тілінің оқу ішкорпусын әзірлеуге қажетті алғышарттар анықталады. Метабелгіленімдері мен тіл үйретуге/үйренуге бағытталған лингвистикалық белгіленімдері, анықтамалық құралдарды тіл үйренушінің қазақ тілін меңгеруіне қарай бейімдеу жолдары, компьютерлік-бағдарламалық әдістері қарастырылады.

Қазақ тілінің лексикалық қабатын оқу корпусына бейімдеу: семантизация және автоматтандыру жолдары. Қазақ тілінің ұлттық корпусы – қазақ сөзінің цифрландырылған күрделі нұсқасы. Себебі қазақ тілінің корпусы тілтанымдық қызметтермен қамтамасыз ете алады, соның негізінде бірнеше ішкорпустардан тұрады. Оның ішкорпустарының ішіне оқу корпусына деген сұраныс күннен күнге артып келеді. Оның себебі Қазақстан көпұлтты мемлекет болғандықтан, мемлекеттік тілден тыс, орыс, ағылшын тілдерінің қолданысы да басымдыққа ие. Сондықтан қазақ мәдениетін тұтынушы өзге ұлт өкілдері өз мәдениетінен пара-пар нұсқасын тану үшін аударма-баламасын анықтағысы келеді. Сондай-ақ оқу корпусы тіл үйренушілерге тиімді тілтанымдық база болары да сөзсіз.

Зерттеу тақырыбында оқу корпусына енетін тілтанымдық қорды, әсіресе қазақ тілінің лексикалық қабатындағы сөздерді тақырыптық, мағыналық, мәдени-семантикалық топтарға іріктеп семантизациялау және цифрлық жүйеге сәйкес автоматтандыруға лайықтау әрі икемдеу, оқу корпусының басқа ішкорпусстардан айырмашылығы туралы мәселе қарастырылып, семантизациялау жіктелімі және мағыналық топтарды кестелеу (автоматтандыруға лайықтау) жолдары ұсынылады.

7. Диалектологиялық ішкорпус. «Қазақ тілі ұлттық корпусының диалектологиялық ішкорпусын әзірлеу технологиясы» тақырыбында диалектологиялық корпус жасаудың әлемдік тәжірибесіне шолу жасалады. Орыс тілінің ұлттық корпусында диалектілік ішкорпус ауызша тіл материалдары бойынша енгізілгендігі, ондағы фонетикалық транскрипция мәселелері мен оларды орфографиялау, просодикалық белгіленімдер әзірлемесі және диалектілік қолданыстарға автоматты морфологиялық талдаулар жасаудың жолдары көрсетіледі. Сонымен қатар Саратов диалектілік корпусындағы метамәтіндік белгіленімдер қоюдың технологиясы сипатталады. Осы диалектілік корпус жасаудың әлемдік тәжірибесі негізінде қазақ тілінің диалектологиялық ішкорпусының әзірлемесі сипатталып, оны жетілдірудің жолдары қарастырылады. Қазақ тілінің диалектілік ішкорпусы қазіргі уақытта «Қазақ тілінің аймақтық сөздігі» материалдары бойынша жасалып отыр. Осы ретте мақалада «Аймақтық сөздік» базасын Диалектілік ішкорпусқа енгізу барысында жасалған өңдеу технологиясы беріледі.

Қазақ тілінің диалектілік ішкорпусында метабелгіленімдер өңірлік, морфологиялық және лексикалық деңгейде белгіленді. Олар: (лексеманың) сөз табы, өңірі, варианты, лексикалық мағынасы, грамматикалық түрленімі, фонетикасы мен фонологиясы. Келешекте Диалектілік ішкорпуста өңірлерден ауызша тіл материалдарын аудиотаспаға жазып, корпусқа енгізу жоспарланған. Осы мақсатқа сәйкес жұмысты ұйымдастыру бойынша ұсыныстар берілді.

8 Мәдени репрезентативті мәтіндер ішкорпусы. Мәдени-семантикалық белгіленім әзірлемесі. Мәдени-репрезентативті ішкорпус қазақ этносының этномәдени даму жолын, рухани және заттық құндылықтарын, этнографиялық, мәдени-танымдық ұғымдар түп-тамырын тереңірек танудың бір құралы ретінде ғана емес, тілдік қазынаның елді тұтастырушылық қызметін сипаттайтын маңызды таным тетігі ретінде қарастырылды. Қазақ тілінің Ұлттық корпусына Мәдени-репрезентативті ішкорпусты енгізу тәжірибесі алғаш рет қолға алынып отыр. Мәдени-семантикалық белгіленімнің негізгі мақсаты мәдени мағынасы жасырын тұрған тілдік бірліктердің астарындағы білімдерді анықтауға негізделеді. Сөздің лексикалық мағынасымен бірге этномаркерлі мәнділіктерін, этномәдени семантикасын, когнитивтік білімдер жүйесін, сөз астарындағы прецедентті мазмұндағы білімдердің бірге берілуімен құнды.

Тақырыптың ғылыми маңыздылығы – ҚТҰК құрамына Лингвомәдени ішкорпусын енгізудің алғышарттары мен этнотанбаланған тілдік бірліктерде жасырын тұрған мәдени кодтарды интерпретациялаудың үлгілерінің көрсетілуі болса, практикалық маңыздылығы – қолданушыға ҚТҰК құрамындағы Лингвомәдени ішкорпусының ерекшеліктері, қолданушыға қажетті ақпарат

алу тәртібінің нұсқаулығы ретінде қосымша бола алады. Аталған ішкорпуста мәдени тілдік бірліктер қамтылып, олардағы кодқа салынған тілдік-мәдени ақпараттарды ашудың тәсілдері көрсетілді.

Этномәдени бірліктердің корпуста берілу жолдары. Бұл тақырыпта қазақ тіліндегі этномәдени бірліктер жүйесіндегі заттық мәдениет атауларына жасалған сипаттамаларды Қазақ тілінің ұлттық корпусына енгізу және корпусты белгілеу, берілу жолдары айтылады.

Этномәдени бірліктер қазақ тілінің ұлттық болмысын танытатын, ұлт тілін, дәстүрін, дүниетанымын, рухани әлемін сақтаушы кумулятивтік қызметі зор тілдік танбалар болып табылады. Осы ретте бүгінгі ұрпаққа тілде бейнеленген этномәдени бірліктер арқылы дүниенің қазақи бейнесін жеткізу – қазіргі жаһандану заманында жас ұрпақтың білім-білігін арттыру үшін оңтайлы да тиімді құрал ретінде өзекті мәселе. Бүгінгі жас ұрпақтың білімді кітаптан емес, ақпараттық құрылғылар арқылы алуды дағдыға айналдырып бара жатқаны белгілі. Осыған орай тіл арқылы бүгінгі ұрпақты ұлттық құндылықтарымызбен сусындатып, ұлттық білім қорын келешек ұрпаққа жеткізу – аса маңызды.

Қазақ тілінің ұлттық корпусы осындай ұлттық мәдени бірліктерді кең арнада сақтаушы, ұрпаққа танытушы қызметін атқаратын заманауи инновациялық-ақпараттық құрал болып табылады.

Лингвомәдени білімдер жүйесін ұлттық корпуста енгізуде көне атақ-лауазым атауларын зерттеу. Белгілі бір халықтың дүниетанымын, ой-санасын, өркениеті мен мәдениетін тіл арқылы, оның ақпараты арқылы білу – қазіргі таңда өзекті мәселелердің бірі болып отыр. Байырғы қазақ қоғамындағы атақ-лауазым атауларының тарихи даму барысында қалыптасқан жүйесі мен мәртебесін, деңгейін, этнолингвистикалық сипатын, мағыналарын, дамуының ішкі-сыртқы дереккөздерін зерделеу – зерттеу жұмысының негізгі мақсаты. Сондықтан қазақ тіліндегі байырғы атақ-лауазым атауларының этнолингвистикалық табиғатын ашып берудің, олардың этномәдени мағыналарын анықтаудың мәні ерекше болмақ.

Зерттеліп отырған лексикалық тіл байлығында ел басқару мен билік айтуға, атақ-лауазымға байланысты мәдени ақпарат қана емес, онда қазақ халқының мәдениеті, менталитеті, көзқарасы мен дүниетанымы, өзіндік болмысын көрсететін құнды деректер сақталған, сондықтан олар – тарих пен уақыттың табын бойына сақтап келген рухани шежірелік сипатымен құнды мұралар; атақ-лауазым атаулары қазақ этносы қалыптасуындағы мәдениет жалғастығын көрсетіп, қазақ қоғамының әр дәуірдегі қалпын, қалыптасу, даму үрдісін сипаттайды.

Мәдени-репрезентативті ішкорпустарды әзірлеудегі мәдени бірліктерді аннотациялау проблемалары. Қазіргі қазақ қоғамындағы аса өзекті мәселелердің бірі – тіл арқылы қалыптасып, тілде сақталған төл мәдениет пен тарих сабақтастығын кешенді зерттеу арқылы халықтың рухын жандандыру. Бұл мәселенің шешімі тілді зерттеудің құрылымдық жүйесінің қазіргі қазақ тіл білімінде бүгінгі таңда кең ауқымды дамып келе жатқан тілдік санадағы дүниенің ұлттық көрінісін бейнелейтін антропоөзекті бағытқа жалғасумен тікелей байланысты. Осы орайда соңғы кезде жүргізілген зерттеулер мен жарық

көрген лексикографиялық еңбектер әр халықтың бар болмысын көрсететін мәдени-репрезентативті мәтіндердің мол ақпаратын пайдаланудың өзектілігін көрсетіп отыр. Өйткені әр ұлттың даралық ерекшеліктері мен бірегейлігі көрініс тапқан мұндай мәтіндер – ұлт қалыптасуының басты көрсеткіші ретінде сол ұлттың коды, баға жетпес асыл мұрасы, тілдік қазынасы. Сондықтан мәдени мәтіндерді жинап, саралап, мағыналарын жете түсіну, қайта жаңғырту – өте өзекті мәселе. Ендеше, көне мәдени мәтіндерімізді түгендеп алып, ұлттық тіліміздің ішкорпусын жасасақ, ұлттық құндылықтарымыздың тілдегі заманауи көрінісі жасалады, кейбір ұмыт болған немесе мағынасы күңгірттенген көнерген сөздер қайта жаңғырып, сөздік қорымыздан қайта орын алады.

Қытайда тұратын қазақтардың тілінде сақталған этномәдени бірліктер тақырыбында Қытайда тұратын қазақтардың тілінде сақталған, отандық сөздіктерге әлі ене қоймаған этномәдени бірліктер жайында айтылады.

«Түйе» сөзінің семантикалық тіркесімділігі және ұлттық корпуста автоматтандыру тәсілдері. Зерттеуде лингвистика ғылымында «түйе» сөзінің тіркесімділік әлеуеті мен семантикалық валенттілігі талданады. Мақаланың мақсаты – «түйе» сөзінің тіркесімділігі жоғары екенін және оған негіз болған концептуалды аялық білімдердің жиынтығын таныту арқылы ішкорпуста рәсімдеу жолдарын ұсыну. «Түйе» сөзіне қатысты автоматтандырылған мәтінді рәсімдеу де іздеу жүйесі арқылы жүзеге асады. Ол мәтіндерді «Түйе» ішкорпусына байлау арқылы, түйе тіркесімі қатысқан сөйлемде семантикалық белгіленімге байлауға әбден болады.

9. Мақал-мәтелдер ішкорпусы: мәтіндік базасы, түсіндірме беру әдістері. Мақал-мәтелдер – әр ұлт тілінде ғасырлар бойы қалыптасып, қауым жадында сақталып, ұрпақтан ұрпаққа рухани мұра болып келе жатқан асыл сөз қазынасы, халық даналығының қорланып жиналған қоры, рухани мұрасы. Халықтың өмір тәжірибесінде еленіп-екшеліп, тұжырымдалған пәлсапа болғандықтан, ондағы тұжырымды ойлар адамзат болмысына, іс-әрекетіне жаппай тән болып келеді. Әсіресе өмірлік тәжірибесі аз жастардың бойына мақал-мәтелдерді сіңіре беру болашақ көргенді ұрпақ тәрбиелеудің бірден-бір құралы екендігі даусыз. Қазіргі жаһандану заманында жастарға, өскелең ұрпаққа мақал-мәтелдерді меңгертудің, саналарына сіңірудің бір жолы – ақпараттық технология жетістіктерін пайдаланып, жастарға, қалың көпшілікке қолжетімді ету. Ал қазіргі кезде қолжетімділік дегеніміз – ол ұялы телефондарда, әлеуметтік желілерде мақал-мәтелдердің базасын салу, пайдаланушыға ыңғайлы етіп Мақал-мәтелдер корпусын жасау болып табылады. Мақал-мәтелдерді әртүрлі мақсатта пайдаланушы адам корпустан өзіне қажетті сөз қамтылған не тақырыптық топқа қатысты мақал-мәтелдерді іздеу ұяшықтарына жазу арқылы тез тауып алуына болады.

Белсенді қолданыстағы паремиологизмдер. Зерттеу авторлары қазіргі кезде белсенді қолданыста жүрген мақал-мәтелдерді анықтауды, олардың қазіргі семантикасын, мәдени-семантикалық мазмұнын ашуды мақсат етеді. Нәтижесінде белсенді қолданыстағы мақалдардың барлығының дерлік бастапқы мағынасы, мәні сақталғанын, алайда, қазіргі қазақ қоғамында белсенді қолданысқа ие паремиологизмдердің арасында көшпелі қазақтың тұрмысында ерекше орын

алған төрт түлік мал жөніндегі, көші-қон жөніндегі, қару-жараққа қатысты т.б. осы күнгі тіршілігімізде жоқ мақал-мәтелдер пассив қорға қарай өтіп бара жатыр деген қорытындыға келеді. Паремнологизмдердің этностың этникалық сипатының, мәдениетінің, мінез-құлқының, ішкі жан дүниесінің, халықтың ғасырлар бойы көңілге түйгенінің сұрыпталған түйіні екеніне тоқталады.

10. Жарнама ішкорпусы. «Қазақ» газетіндегі жарнама мәтіндерінің құрылымы. Зерттеудің негізгі мақсаты – Алаш зиялылары ұйымдастырып, 1913-1918 жылдар аралығында шығарып тұрған алғашқы ұлттық баспасөз – «Қазақ» газетіндегі жарнама мәтіндерін тақырыптық топтарға жіктей отырып, тілдік ерекшелігіне талдау жасау болды. Ол үшін алдымен аталған газеттің 1913 жылғы 7-40 сандарындағы жарнама мәтіндері төте жазудан, транскрипцияланып, графикалық проекциясы жасалды. Зерттеуде қазақ жарнамасының тарихына да шолу жүргізілді. Қазақ тіліндегі тарихи жарнама мәтіндерінің сөйлем құрылысы, сөз тіркестіру амалы, сөзсаптамы талдауға алынды. «Қазақ» газетінің 1913 жылғы 7-40 сандарында жарияланған тарихи қазақ жарнама мәтіндері негізінде ХХ ғасырдың басындағы жарнама түрлері анықталды. Мақалада «Қазақ» газетінде сауда және әлеуметтік жарнамалар жиі жарияланғаны анықталып, ХХ ғасырдың басындағы жарнаманы безендіру технологиясының ерекшеліктері жан-жақты баяндалды.

Жарнамада берілген имплицитті мәтіндерді түсіну ерекшелігі. Жарнама мәтіндерін түсіну кезінде тіл иелменінің санасында танымдық үдерістер мен механизмдер іске қосылады. Сондықтан жарнаманың санада көрініс табуына қатысты лингвистикалық мәселелерді шешу үшін таным үдерісі мен оның механизмдеріне жүгінеміз. Жарнама мәтінін түсіну оның анық берілген мағыналарын оқумен ғана шектелмейді, сондай-ақ инференция үдерісі арқылы жүзеге асатын астыртын мағыналарды ажыратумен де жүзеге асырылады. Сондықтан мақалада жарнама коммуникациясындағы мәтіннің инференциялану үдерісі де қарастырылады.

11. Лингвистикалық белгіленім әзірлемелері. Дерексіз зат есімдер: Қазақ тілінің ұлттық корпусында лексикалық белгіленімді анықтаудың кейбір мәселелері. Дерексіз зат есімдер – қазақ тілтанымында және түркологияда аз зерттелген лексикалық қабаттың бірі. Дерексіздік категориясына енетін зат есімдер – абстрактілі ұғымдарды, қасиет пен сапаны, іс-әрекет пен жай-күйдің әртүрлі деңгейін танытатын, қазақ тілінің сөздік құрамындағы ауқымды қабат. Дерексіз зат есімдердің басты ерекшелігі сонымен қатар тілді қолданушының қоршаған орта мен ақиқат дүниеге қарым-қатысын білдіретіндігіне байланысты.

Бұл тақырыпта Қазақ тілінің Ұлттық корпусына тіркелетін дерексіз зат есімдердің лексикалық-семантикалық белгіленімдерін анықтау, ажырату жолдары бойынша ұсыныстар жасалады.

Ахмет Байтұрсынұлының және өзге де белгілі ғалымдар мен түркологтардың жіктелімдері негізінде қазақ тіліндегі дерексіз зат есімдер құрылымдық, мағыналық жағынан топтастырылып, оларды тану, дерексіз зат есімдердің мағыналық ерекшеліктері, олардың лексикалық-семантикалық белгіленімдерін анықтау жолдары қарастырылады.

Зат есімдердің лексика-семантикалық белгіленімін ұлттық корпуста рәсімдеу жолдары. Зат есімнің лексика-семантикалық белгіленімі – корпуста енгізілген мәтін бойындағы зат есімдердің лексика-семантикалық мәліметінен хабар беретін таңбалық жүйе. Қазақ тілінің ұлттық корпусында зат есімнің лексика-семантикалық белгіленімін рәсімдеуде «Формадан – мағынаға қарай» қағидасы негізге алынды. Зат есімнің лексика-семантикалық белгіленімінің жіктелімі: реті, морфологиялық белгіленімі – сөз табы – (зат е., сын е); сөз-жасамдық белгіленімі (Сөздің морфемалық құрамы (түбір; түбір+қосымша; түбір+түбір)); сөзжасамдық тәсілі (синтетикалық/аналитикалық/лексика-семантикалық); тұлғасы (негізгі/туынды); құрамы (дара/ күрделі: қос сөз, біріккен, кіріккен); морфо-семантикалық белгіленімі (нақтылық/ абстрактылық сипаты (деректі/дерексіз)); адамзаттық/ғаламзаттық сипаты (жанды/жансыз); жалпы/жалқылық сипаты (жалпы/жалқы); лексика-семантикалық белгіленімі: (лексикалық қабаты (төл/кірме/шетілдік/жерг/кәсіби/тарихи/көне/жаңа/ кітаби); стильдік реңкі (сөйлеу/ қарапайым/экспр/эмоц, қарғыс/алғыс/ эвфемизм); терминдік қолданысы (спорт, өнер, муз, хим, мед, психол, авиац. анат. аңшылық, археол. астр. ауылшар., әдеб. әск. биол. бот. геолог. геогр. геом. діни, заң, зерг. зоол. т.б. бот.); тақырыптық-семантикалық тобы (макротоп, микротоп) бойынша белгіленді.

Қазақ тілінің ұлттық корпусындағы етістіктердің лексика-семантикалық белгіленімі әзірлемесі: әлемдік тәжірибе, классификациялау, корпуста белгілеу. Қазақ тілінің Ұлттық корпусын жасау тілдің әр деңгейлері бойынша автоматты түрде талдау жасайтын белгіленімдер жасауға негізделді. Сөздерді лингвистикалық аннотациялаудағы күрделі белгіленімнің бірі – лексика-семантикалық белгіленім. Бұл тақырыпта әлемдік корпус жасау тәжірибесіндегі негізгі белгіленімнің бірі лексика-семантикалық белгіленім әзірлемесін жасау мәселесі қарастырылады. Нақтырақ айтқанда, компьютерлік лингвистикаға және лексика-семантикалық классификацияға қатысты отандық және шетелдік ғалымдар еңбектеріне шолу жасалады, қазақ тілінің Ұлттық Корпусындағы етістіктердің лексика-семантикалық белгіленімін жасау кезеңдері көрсетіледі, практикалық негізі түсіндіріледі.

Орыс, қалмақ және т.б. тілдер корпусымен салыстырғанда Қазақ тілінің Ұлттық корпусындағы лексика-семантикалық белгіленім сөз мағынасына, яғни семаға тереңірек бойлайды.

Қазақ тілінің ұлттық корпусында етістіктің лексика-семантикалық белгіленімін қою жолдары атты тақырыпта етістіктердің лексика-семантикалық, лексика-грамматикалық белгі-кодтары анықталып ірі және ұсақ мағыналық топтарға жіктеу жұмыстары жөнінде айтылады. Етістіктерге лексика-семантикалық белгіленім қою арқылы қалай автоматтандырылады деген мәселе төңірегінде сөз болады. Етістіктің лексика-семантикалық іш-корпусын дайындау үдерісі ең алдымен, жинақталған етістіктердің семантикалық ерекшелігіне қарай макро мен микро топтарға бөлуден басталады. Одан кейін барып лексика-грамматикалық, морфологиялық, сөзжасамдық, коннотациялық белгіленімдер әзірлемесін дайындаудың жолдары көрсетіледі.

Лексика-семантикалық белгіленім бағдарламасын жасауға қажетті етістіктің мағыналық топтарының классификациясы жасалып, сол арқылы Институтта етістік сөз табының лексика-семантикалық белгіленімдерін ішкорпусқа енгізу жолдары қарастырылады.

Корпуста күрделі етістіктерді автоматты тану және лексикографиялау мәселелері. Табиғи тілді өңдеуге арналған арнайы компьютерлік бағдарламалардың ішінен мәтінге белгіленім енгізуді автоматты түрде жүзеге асыратын бағдарламалар айырықша орын алады. Морфологиялық талдауды (тэггерлер) автоматтандыратын компьютерлік бағдарламалар жұмысының нәтижесінде мәтін бойындағы әрбір лексикалық бірлікке олардың грамматикалық сипаттамалары «тіркеліп» беріледі. Корпуста морфологиялық талдауыш тіркесті күрделі сөздерді біртұтас бірлік ретінде танымайды, жеке сөздерге бөліп талдайды. Зерттеуімізде күрделі етістікті тіркестерді автоматты морфологиялық талдауда біртұтас лексика-грамматикалық бірлік ретінде танудың жолдары қарастырылады. Күрделі етістіктер сөздігін жасау қажеттігі негізделеді.

ҚТҰК-дағы сөзжасамдық белгіленім әзірлемесі тақырыбында Қазақ тіліндегі лингвистикалық белгіленімдердің бірі – сөзжасамдық белгіленімдерді корпусқа енгізудің әдіс-тәсілдері қарастырылады. Мәтінді автоматты тану формальді құрылымға негізделетіндіктен, сөзжасамның синтетикалық және аналитикалық тәсілдері арқылы жасалған туынды сөздерді (қос сөз, біріккен сөз, қосымшалы сөз) кодтау технологиясы көрсетіледі.

КРАТКАЯ АННОТАЦИЯ

Монография состоит из 11 разделов, посвященных опыту создания подкорпусов казахского языка.

1. Концептуальные основы создания Национального корпуса казахского языка

В данном разделе описываются концептуальные основы создания Национального корпуса казахского языка, его настоящее и будущее, а также этапы его разработки в Институте языкознания им. А. Байтурсынова.

2. Устный подкорпус. Поскольку устная речь раскрывает и позволяет познать языковые традиции определенной нации, подобные тексты обогащают содержание национального корпуса. Обоснована необходимость включения в подкорпус устной речи следующих текстов: разговорно-бытовой стиль (диалоги между членами семьи, родственниками и знакомыми, соседями; разговоры в магазинах, транспорте, общественных местах; неформальные разговоры по телефону), неформальная беседа, традиционная, импровизированная речь в интервью, устные выступления, заранее подготовленные речи на лекциях, конференциях, круглых столах, советах, семинарах и интервью, дискурс кино и сценических представлений, заранее подготовленный и импровизированный устный религиозный дискурс, импровизированная речь на теле-радиопрограммах с известной тематикой. Импровизированная речь является истинной и природной. В связи с этим в статье описываются национальные особенности устной речи. В частности, анализируется традиционная речь пожилых людей из местностей с преимущественно казахским населением.

Устный подкорпус казахского языка: разработка фонетической, просодической разметки в корпусе текстов. В процессе исследования были накоплены и систематизированы знания, связанные с сегментной и суперсегментной фонетикой казахского языка, а также была подготовлена просодическая разметка. Предоставлены фонетические, просодические описания каждого исконно казахского звука, расставлены условные обозначения интонационных компонентов. Рассмотрены пути фонетической и просодической разметки текстов.

Предоставлена просодическая разметка текстов в базе устного подкорпуса, создана орфоэпическая норма речи говорящего. Данный подкорпус послужит фундаментом для многочисленных исследований, посвященных фонетико-просодическим особенностям слова.

Фонетико-фонологическая и интонационная транскрипция текста в базе устного подкорпуса

В исследовании рассматривается роль фонетико-фонологической транскрипции в современной лингвистике и сфера ее использования. В современной лингвистике уделяется мало внимания просодическому аспекту транскрипции текстов. Исследование опирается на труды академика З.М. Базарбаевой, впервые детально рассмотревшей проблему интонации в казахской лингвистике. В процессе анализа текста рассмотрены модели восьми основных интоном. Иными словами, помимо фонетической транскрипции будет предоставлена и интонационная транскрипция текста с целью определения просодических осо-

бенностей. Также при транскрипции текста каждое предложение членится на синтагмы и ритмические группы; отмечаются особыми знаками и описываются такие компоненты, как тембр, мелодия, пауза, темп. Даются модели интоном каждого предложения.

Опыт создания орфоэпического конвертера

Конвертер – это специальная программа, преобразовывающая файлы одного типа в файлы другого типа. Орфоэпический конвертер – это программа, которая переписывает изначально написанные по правилам орфографии слова, словосочетания или тексты согласно нормам орфоэпии.

Отделы фонетики и прикладной лингвистики Института были ответственны за создание орфоэпического конвертера. Проанализировав существующие труды по фонетике, отделы подготовили дополненный алгоритм орфоэпических правил. Алгоритм, состоящий из 183 правил, был подкреплён примерами.

Подготовленный конвертер способен решить несколько важных проблем. Во-первых, он дополняет внутреннюю разметку Национального корпуса казахского языка (<https://qazcorpus.kz/>) и верно делит слова на слоги. Во-вторых, он переписывает тексты из базы устного подкорпуса, руководствуясь нормами орфоэпии. В-третьих, конвертер будет выпущен в виде отдельного продукта. Для достижения качественного результата при создании речевых технологий необходимо наличие автоматизированных орфоэпических норм и словарей.

3. Ономастический подкорпус

Институт языкознания им. А. Байтурсынова начал разработку Ономастического подкорпуса, который является независимым компонентом Национального корпуса казахского языка. В рамках данной темы обсуждается направление создания Ономастического подкорпуса и его принципы. В результате выявляются основные критерии создания Ономастического подкорпуса. Описываются как Национальный корпус казахского языка, так и опыт крупных мировых корпусов. Проводятся сравнения, связанные с разработкой топонимической разметки, выдвигаются предложения. Как можно заметить из опыта зарубежных корпусов, наличие большого количества подкорпусов только обогащает национальный корпус. Появление Ономастического подкорпуса создает множество новых возможностей для исследований ученых-ономастов, поскольку ускоряется поиск текстов, где встречаются географические названия. Это не только упрощает процесс поиска, но и вносит вклад в комплексные исследования в сфере ономастики. Ономастический подкорпус, безусловно, необходим не только с точки зрения утверждения и стандартизации географических названий, но и для систематизации топонимов. Ономастический подкорпус позволяет сохранить национальный фонд топонимов и предоставляет доступ к их научному анализу, удовлетворяя познавательный интерес.

Технология введения мифонимов и теонимов в Национальный корпус казахского языка

В ономастической отрасли казахской лингвистики были проведены определенные исследования, защищены диссертационные работы, составлены словари наименований. Однако не наблюдается последовательность в исследованиях, посвященных раскрытию смысла мировоззренческих взглядов,

происходящих из познания древними этносами картины окружающего мира, которая, в свою очередь, хранится в наименованиях и основывается на мифах, мифонимах и теонимах. В связи с этим в рамках данной темы сравниваются материалы, посвященные технологиям включения в корпус мифологических и религиозных наименований путем тщательно проанализированной лингвистической разметки, которая находится в открытом доступе в сети Интернет. Анализируются общие воззрения и концепции, связанные с наименованиями соответствующих мифонимов и теонимов. Дается информация о технологии включения мифонимов и теонимов в Национальный корпус казахского языка. Данный процесс позволяет модернизировать мифологическое и религиозное пространство казахского языка через систему духовных ценностей, а также способствует повышению эффективности процесса обучения языку.

4. Из мирового опыта создания исторических подкорпусов

Понятие Национального корпуса является инструментом не только синхронических, но и диахронических исследований. В связи с этим упоминается необходимость создания «исторического подкорпуса» в рамках Национального корпуса казахского языка. Исторический подкорпус весьма полезен и необходим при исследовании языка рукописных памятников V-XX веков, а также истории, культуры и литературы данного периода. В области прикладной лингвистики такой Исторический подкорпус разрабатывается впервые. Однако существуют трудности оцифровки рукописей, дошедших до нас в различной графике. Исследование фокусируется на структуру и практическое применение разработки исторического подкорпуса в мировом опыте. Учитывается опыт других стран при определении основных работ, которые будут проведены при разработке исторического подкорпуса. Рассматриваются такие вопросы, как определение и сбор базы текстов рукописных памятников отдельных периодов; селекция полученных материалов по качеству и составу, классификация, обработка; проблемы введения текстов в корпус и их демонстрации; структура метаразметки каждого рукописного памятника. Предлагаемое исследование может быть использовано при разработке исторического подкорпуса любого языка.

Опыт Национального исторического корпуса русского языка и представление тюркского лексического слоя

В настоящее время бурно развиваются связи в области языка и искусственного интеллекта, языковой и компьютерной прикладной лингвистики. Продолжением и новой вехой этой связи является требование сегодняшнего дня рассмотреть рукописные памятники языка как объект компьютерной прикладной лингвистики. В прикладной лингвистике современного мира сформирован опыт создания корпусов, в том числе исторических, что расширяет возможности для многих научных исследований. С этой целью был изучен Исторический корпус русского языка (прикладная лингвистика и построение корпусов которого хорошо развиты) и рассмотрена практика передачи архаических языковых единиц и проявления в ней тюркского лексического пласта. Кроме того, был рассмотрен национальный корпус армянского языка с целью оценки опыта создания всемирно-исторического корпуса. В ходе создания и совершенствования исторического корпуса казахского языка мы предлагаем возможность

создания корпусов исторических словарей, исторической ономастики, исторической терминологии, исторической орфографии, исторической грамматики, исторического словообразования, исторической фразеологии, исторической сочетаемости, исторического богословия, старых религиозных текстов и исторической частотности лексем. Создание Исторического корпуса является одним из механизмов предотвращения поглощения любого языка глобализацией. Исторический корпус — это современный цифровой мост между библиотекой и читателем.

Лексико-грамматическая разметка текстов Исторического подкорпуса. Лингвистическая разметка – это лингвистическая информация, предоставляемая для каждой единицы текста в орфографическом, фонетическом, лексическом и грамматическом аспектах. Однако подготовка лингвистической разметки текстов Исторического подкорпуса – это чрезвычайно сложная работа, требующая глубоких изысканий. Так, основная масса текстов, введенных в Исторический подкорпус, написана арабской графикой. Эти тексты по-разному транскрибированы в кириллицу. Разработка Исторического подкорпуса, по сравнению с другими подкорпусами, сопровождается большими теоретическими и техническими трудностями. В связи с этим была рассмотрена проблема лингвистической разметки текстов Исторического подкорпуса. В этом плане рассматривается опыт разработки исторического подкорпуса русского языка; анализируется транскрипция текстов, написанных арабской графикой в разные периоды; определяются варианты транскрибированных слов; уточняется лексико-грамматическая разметка транскрибированных текстов; описываются механизмы лексико-грамматического поиска по транскрибированным текстам.

5. Параллельный подкорпус. Основные понятия переводоведения

Известно, что истоки такой актуальной, полезной и важной отрасли, как переводоведение или наука о переводе, кроются в истории какого бы то ни было народа. В культуре каждой нации, образовавшей народ, перевод становится неотъемлемой частью общественной жизни – в особенности после появления письменности; высказываются предположения и мнения о переводе. В подобных предположениях часто заметен поиск ответов на вопросы о том, какова роль перевода в обществе и в культурной жизни, какой вклад он вносит в культуру, в чем его особенности. Понятие лингвистической теории перевода сформировалось в двадцатых годах 20-го столетия: изначально этот термин рассматривался в тесной связи с понятием художественного перевода.

«Параллельный подкорпус казахского и русского языков: разработка, использование и проблемы». В рамках данной темы дается краткое описание видов параллельных подкорпусов. В параллельном подкорпусе, который разрабатывается в рамках Национального корпуса казахского языка, вручную сопоставлены оригинальный текст романа-эпопеи «Путь Абая» М. Ауэзова и его перевод на русский язык, созданный А. Ким.

В процессе разработки Параллельного подкорпуса были использованы программные инструменты Microsoft Office Excel, Notepad++, Python, Django, MySQL.

Параллельный подкорпус состоит из *базы сопоставляемых текстов, системы поиска и информации о введенных текстах (метатарметки): автор, переводчик, название произведения, название перевода, время публикации произведения, время перевода, язык оригинала, язык перевода*. Поисковая система позволяет найти искомое слово по таким параметрам, как *слово, словосочетание, предложение, заглавная буква* (в казахском и русском языках).

Особенности перевода невербальных средств в романе «Путь Абая».

Роман-эпопея «Путь Абая» – это художественное произведение высокой культурно-языковой ценности, предоставляющее обширную информацию об истории и быте казахского народа. Начиная со второй половины XX века по настоящее время роман постоянно переводится на множество языков мира. Однако эти переводы осуществляются не с оригинала, т.е. с казахского языка, а посредством языка-донора – русского. Спорным вопросом является сохранение в таком опосредованном переводе не только оригинальной идеи автора, но и художественного влияния текста и его языковых, познавательных особенностей. В связи с этим особенно актуален лингвистический анализ перевода языковых единиц в романе «Путь Абая». Невербальные средства, дающие возможность познать национальные особенности народа, также играют важную роль в любом художественном тексте с точки зрения и количества, и качества, т.е. в аспекте экспрессивного, информативного значения. В исследовании проведен сравнительный анализ параэлементов в оригинальном тексте и их эквивалентов в переводном тексте. Так, выявлено и подкреплено примерами, что жесты переведены шестью следующими способами: *непереводимые языковые единицы (лакуны); единицы, переведенные дословно; единицы, переведенные качественно; сопоставленные единицы; свободно переведенные единицы и, наконец, наполовину переведенные языковые единицы*.

Лексикографический потенциал параллельного корпуса текстов.

Данная тема посвящена исследованию лексикографического потенциала Параллельного корпуса текстов. В работе уделяется особое внимание методике и инструментам анализа, раздельной демонстрации оригинальных данных в параллельных корпусах в лексикографическом аспекте. Результаты обогащают наше понимание лексикографии и раскрывают перспективы современных исследований в области разработки параллельных корпусов и компьютерной обработки естественного языка.

6. Учебный корпус. Вопросы разработки учебного корпуса казахского языка. В рамках данной темы проводится обзор мирового опыта разработки учебных корпусов. Исследуются такие вопросы, как типы текстов, охваченных в Учебных корпусах английского и русского языков, виды лингвистической информации, предоставленной для изучающих язык и т.д. Также определяются предпосылки разработки учебного корпуса казахского языка Институтом языкознания им. А. Байтурсынова. Рассматриваются метатарметка и лингвистическая разметка, направленная на изучение языка/обучение языку, пути и компьютерно-программные методы адаптации справочных материалов к усвоению казахского языка человеком, изучающим язык.

Адаптация лексических слоев казахского языка к учебному корпусу – пути автоматизации и семантизации

Национальный корпус казахского языка – это сложный оцифрованный вариант казахской речи. Корпус казахского языка, состоящий из нескольких подкорпусов, способен предоставить возможность познания языка. В частности, возрастает спрос на Учебный корпус. Причина кроется в том, что в Казахстане – многонациональном государстве – широко употребляются не только государственный язык, но и русский, и английский. Поэтому представители других наций изъявляют желание найти эквиваленты элементов казахской культуры в своей культуре и обращаются к переводам. Также учебный корпус, несомненно, станет эффективной познавательной-языковой базой для тех, кто изучает язык.

В исследовании рассматриваются семантизация и деление познавательного-языкового фонда, вводимого в Учебный корпус, на тематические, семантические, культурно-семантические группы, адаптация и приспособление его к автоматизации, соответствующей цифровой системе, а также проблема отличия Учебного корпуса от других подкорпусов; предлагаются пути семантической классификации и отображения групп по значению (адаптация к автоматизации).

7. Диалектологический подкорпус. Технология разработки диалектологического подкорпуса казахского языка

В рамках данной темы проводится обзор мирового опыта создания диалектологических корпусов. В Национальном корпусе русского языка диалектологический подкорпус составлен из материалов устной речи. В исследовании показаны проблемы фонетической транскрипции данных материалов, их орфографизация, разработка просодической разметки и пути разработки автоматических морфологических разборов диалектологических употреблений. Помимо этого, описывается технология внедрения метатекстовой разметки в Саратовском диалектологическом корпусе. На основе этого опыта создания диалектологических корпусов описывается разработка Диалектологического подкорпуса казахского языка и рассматриваются пути его улучшения. В настоящее время Диалектологический корпус казахского языка разрабатывается на материале «Регионального словаря казахского языка». В связи с этим в статье дается технология обработки базы «Регионального словаря» в процессе введения ее в подкорпус.

Метаразметка Диалектологического корпуса казахского языка дана на региональном, морфологическом и лексическом уровнях. В частности: часть речи (лексемы) слова, регион, вариант, лексическое значение, грамматическое преобразование, фонетика и фонология. В будущем планируется записать на аудиокассету и ввести в подкорпус материалы региональной устной речи. В рамках этой цели даны предложения по организации работы.

8 Подкорпус культурно-репрезентативных текстов. Разработка культурно-семантической разметки

Лингвокультурный подкорпус рассмотрен не только как способ глубинного ознакомления с этнокультурным развитием казахского этноса, его духов-

ными и материальными ценностями и истоками этнографических и культурно-познавательных понятий, но также как значимый инструмент познания, описывающий национально-объединяющую функцию языкового наследия. Впервые предпринимаются работы по введению Лингвокультурного подкорпуса в Национальный корпус казахского языка. Основная цель культурно-семантической разметки – определение знаний, скрытых за языковыми единицами, культурное значение которых предстоит раскрыть. Подкорпус ценен тем, что в нем помимо лексического значения слова даются его этномаркированные значимости, этнокультурная семантика, когнитивная система знаний и знания прецедентного содержания, скрытые в слове.

Научная значимость темы – демонстрация предпосылок внедрения Культурно-репрезентативного корпуса в НККЯ и моделей интерпретации культурных кодов, скрытых в этномаркированных культурных единицах. Практическая значимость исследования – может быть использовано как справочник по особенностям Культурно-репрезентативного подкорпуса НККЯ и по алгоритму получения нужной информации. Показаны приемы раскрытия культурно-языковой информации, заложенной в коде культурных языковых единиц, охваченных в данном подкорпусе.

Способы представления этнокультурных единиц в корпусе

Данная тема посвящена путям введения в НККЯ описаний материальной культуры в системе этнокультурных единиц казахского языка, а также способам их разметки и демонстрации в корпусе.

Этнокультурные единицы – это языковые знаки с кумулятивной функцией, которые позволяют познать национальную идентичность казахского языка и хранят язык этноса, его традиции, мировоззрение и духовный мир. В связи с этим актуальной проблемой становится передача современному поколению исконно казахской картины мира, отраженной в этнокультурных единицах: эффективный и доступный инструмент улучшения знаний нынешнего поколения в эпоху глобализации. Известно, что современное поколение предпочитает получать знания не из книг, а через технологические средства. Таким образом, чрезвычайно важно обогатить современное поколение нашими национальными ценностями и передать национальный фонд знаний будущему поколению.

Национальный корпус казахского языка является современным инновационно-информационным инструментом, который выполняет функцию сохранения национальных культурных единиц и передачи их будущим поколениям.

Исследование древних наименований титулов и должностей при введении системы лингвокультурных знаний в национальный корпус

На сегодняшний день одной из актуальных проблем является познание мировоззрения, сознания, цивилизации и культуры определенного народа через его языковую информацию. Основная цель исследования – изучить исторически сложившуюся систему, иерархию и уровни наименований титулов и должностей в стародавнем казахском обществе, их этнолингвистический характер, их значения, а также внутренние и внешние истоки их развития. Поэтому особое значение приобретает определение этнокультурного значения

древних наименований титулов и должностей в казахском языке и раскрытие их этнолингвистической природы.

В исследуемом лексическом богатстве не только заключена культурная информация об управлении, власти, титулах и должностях; в ней хранятся ценные данные, показывающие культуру, менталитет, мировоззрение и познание, а также идентичность казахского народа. Поэтому эти наименования – ценное наследие духовно-летописного характера, в котором хранится поступь истории и времени. Наименования титулов и должностей показывают последовательность культуры при формировании казахского этноса и описывают строение, формирование и развитие казахского общества в каждом периоде истории.

Проблемы аннотации культурных единиц при разработке культурно-репрезентативных подкорпусов

Одна из весьма актуальных проблем современного казахского общества – возрождение духа народа через комплексный анализ взаимосвязи исконной культуры и истории, созданной через язык и сохраненной в языке. Решение данной проблемы напрямую связано с соприкосновением структурной системы исследования языка с антропоцентрическим направлением современной казахской лингвистики, которое активно развивается на сегодняшний день и в русле которого отображается национальная картина мира. Проведенные в последнее время исследования и опубликованные лексикографические труды показывают актуальность использования расширенной информации, заключенной в культурно-репрезентативных текстах, показывающих идентичность каждого народа. Такие тексты, в которых реализуются уникальные особенности и аутентичность каждого народа – это национальный код этноса, его бесценное наследие, его языковая сокровищница. Поэтому весьма важной проблемой является сбор, классификация, глубинное понимание и модернизация культурных текстов. Таким образом, собрав древние культурные тексты и создав подкорпус национального языка, мы получим доступ к современному языковому отображению национальных ценностей. Устаревшие слова, чьи значения были забыты или потускнели, будут модернизированы и вновь войдут в лексический фонд.

Этнокультурные единицы, сохраненные в языке казахов, проживающих в Китае. В рамках данной темы рассмотрены этнокультурные единицы, которые встречаются в языке казахов, проживающих в Китае, и которые еще не были включены в отечественные словари.

Семантическая сочетаемость слова «верблюды» и способы его автоматизации в Национальном корпусе

В исследовании анализируется потенциал сочетаемости слова «верблюды» и его семантическая валентность в лингвистическом аспекте. Цель статьи – показать высокую сочетаемость слова «верблюды» и предложить пути его оформления в подкорпусе через представление концептуально-отраслевых знаний, служащих его основой. Оформление автоматизированного текста, связанного со словом «верблюды», также воспроизводится через поисковую систему. Привязав эти тексты к подкорпусу «верблюды», можно внедрить семантическую разметку предложений, в которых встречается данное слово.

9. Подкорпус пословиц и поговорок: текстовая база, методы представления разъяснений. Пословицы и поговорки – это сокровищница слов, фонд народной мудрости, веками формируемый в национальном языке, хранимый в общественном сознании и передаваемый из поколения в поколение в виде духовного наследия. Поскольку это философия, сформированная и концептуализированная в процессе жизненного опыта народа, мысли, представленные в ней, свойственны человеческим действиям и человеческому характеру в целом. В особенности неоспорим тот факт, что внедрение пословиц и поговорок в сознание молодежи, не обладающей обширным жизненным опытом, – это эффективный инструмент создания дальновидного поколения. В эпоху глобализации молодежь и подрастающее поколение можно обучить пословицам посредством достижений информационных технологий, сделав пословицы и поговорки общедоступным достоянием. Доступность в настоящее время – это внедрение базы пословиц и поговорок в сотовые телефоны, социальные сети, это разработка подкорпуса пословиц и поговорок, удобного для использования. Пользователь корпуса может легко отыскать необходимые ему пословицы и поговорки через ключевые слова или тематические группы, вписав их в ячейки поиска.

Активно используемые паремиологизмы

Авторы исследования задаются целью определения пословиц и поговорок, активно используемых в настоящее время, и раскрытия их современной семантики, их культурно-семантического содержания. В результате выявляется, что активно используемые пословицы практически полностью сохранили свое исконное значение и смысл. С другой стороны, пословицы и поговорки, имевшие большое значение для быта кочевых казахов (о скоте, о кочевьях, об оружии), но не относящиеся к современному быту, не представлены среди паремиологизмов активного использования и постепенно переходят в пассивный словарь. Авторы останавливаются на том, что паремиологизмы – это совокупность этнического характера этноса, его поведения, его сознания, а также многовекового опыта народа.

10. Подкорпус рекламы. Структура рекламных текстов в газете «Казах»

Основная цель исследования – классификация по тематическим группам и лингвистический анализ рекламных текстов, представленных в газете «Казах», организованной интеллигенцией Алаш и выпускавшейся в 1913-1918 годах. Для этого, во-первых, были транскрибированы рекламные тексты из 7-40 номеров газеты 1913 года, созданы графические проекции. В исследовании также произведен обзор истории казахской рекламы. Проанализирована структура предложения исторических рекламных текстов на казахском языке, способы сочетания слов, высказывания. На основе исторических рекламных текстов, опубликованных в 7-40 номерах газеты «Казах» в 1913 году, были выявлены виды рекламы в начале двадцатого века. Определено, что в газете «Казах» чаще всего публиковались торговый и общественный виды рекламы. Подробно описаны особенности технологии создания рекламы, применявшиеся в начале двадцатого столетия.

Особенности понимания имплицитных текстов, представленных в рекламе

При понимании текстов рекламы в сознании носителя языка включаются познавательные процессы и механизмы. По этой причине мы обращаемся к процессу познания и его механизмам для решения лингвистических проблем, связанных с отображением рекламы в сознании. Понимание рекламного текста не ограничивается прочтением смыслов, ясно показанных в нем, но также осуществляется путем различения скрытых смыслов, осуществляемых через процесс инференции. Поэтому в статье также рассматривается процесс инференциации текста в рекламной коммуникации.

11. Разработка лингвистической разметки. Абстрактные существительные: некоторые вопросы определения лексической разметки в Национальном корпусе казахского языка

Абстрактные существительные – это лексический пласт, который редко исследуется в казахской лингвистике и в тюркологии. Такие существительные, обозначающие абстрактные понятия, свойства и качества, различные стадии действия и состояния, представляют собой обширный пласт казахского словесного фонда. Основная особенность абстрактных существительных заключается и в том, что они обозначают отношение говорящего к окружающей среде и к действительности.

В рамках данной темы приводятся предложения по определению и разграничению лексико-семантической разметки абстрактных существительных, вводимых в Национальный корпус казахского языка.

На основе классификации Ахмета Байтурсынова и иных известных ученых и тюркологов абстрактные существительные казахского языка сгруппированы по структуре и смыслу, рассмотрены пути определения их лексико-семантической разметки, а также их смысловые особенности.

Пути оформления лексико-семантической разметки существительных в национальном корпусе. Лексико-семантическая разметка существительных – это система знаков, предоставляющая лексико-семантическую информацию о существительных, включенных в корпус. При оформлении данной разметки в Национальном корпусе казахского языка, установка «От формы к значению» была взята за основу. Лексико-семантическая классификация существительных: морфологическое обозначение – часть речи (сущ., прил.); словообразовательное обозначение (морфемный состав слова (корень; корень+аффикс; корень+корень); способ словообразования (синтетический/аналитический/лексико-семантический); форма (основное/производное); состав (простое/сложное; составное); морфо-семантическая разметка (конкретный/абстрактный характер (конкретность/абстрактность)); одушевленный/неодушевленный характер (одушевленность/неодушевленность); нарицательность/собственность (нарицательное/собственное); лексико-семантическая разметка: (лексический слой (исконное/заимствованное/иностранное/местное/профессиональное /историческое/архаическое/новое/ книжное); стилевой оттенок (разговорное/ просторечное/экспрессивное/эмоциональное, проклятие/благодарность/ эвфемизм); терминологическое использование (авиаци., анат.,

археол., астр., биол., бот., воен., геолог., геогр., геом., закон., зоол., искусство, лит., мед., муз., охот., психол., религ., спорт., сельскохозяй., хим., ювел. и пр.); тематико-семантическая группа (макрогруппа, микрогруппа).

Разработка лексико-семантической разметки глаголов в Национальном корпусе казахского языка: мировой опыт, классификация, обозначение в корпусе

Создание национального корпуса казахского языка базируется на разработке разметки, осуществляющей автоматический анализ по каждому уровню языка. Один из сложных видов разметки при лингвистической аннотации слов – это лексико-семантическая разметка. В рамках данной темы рассматривается разработка лексико-семантической разметки как основной разметки в мировом опыте создания корпусов. В частности, проводится обзор работ отечественных и зарубежных исследователей в областях компьютерной лингвистики и лексико-семантической классификации, показываются этапы разработки лексико-семантической разметки глаголов Национального корпуса казахского языка, объясняются практические основы.

По сравнению с корпусами русского, калмыцкого и прочих языков, лексико-семантическая разметка Национального корпуса казахского языка углубляется в значение слова, то есть в сему.

Пути осуществления лексико-семантической разметки в Национальном корпусе казахского языка. В данном разделе говорится о работе, посвященной классификации глаголов на крупные и мелкие семантические группы путем определения их лексико-семантических, лексико-грамматических обозначений-кодов. Рассматривается вопрос о том, как осуществляется автоматизация лексико-семантической разметки глаголов. Процесс подготовки лексико-семантического подкорпуса глаголов начинается, в первую очередь, с разделения подобранных глаголов на макро- и микро-группы с учетом их семантических особенностей. Далее показываются пути разработки лексико-грамматической, морфологической, словообразовательной, коннотативной разметки.

Создается классификация семантических групп глаголов, необходимая для разработки программы лексико-семантической разметки, и рассматриваются пути введения лексико-семантической разметки глаголов в подкорпус.

Проблемы автоматического распознавания и лексикографизации сложных глаголов в корпусе

Среди программ, посвященных обработке естественного языка, особое место занимают программы, осуществляющие автоматическую разметку текста. Компьютерные программы, автоматизирующие морфологический анализ (тэггеры), в процессе работы «присваивают» каждой лексической единице в составе текста ее грамматическое описание. Морфологический анализатор корпуса не распознает сложные слова как цельную единицу, а рассматривает каждое слово отдельно. В исследовании рассмотрены пути распознавания сложных глаголов как цельных лексико-грамматических единиц при автоматическом морфологическом анализе. Обосновывается необходимость создания словаря сложных глаголов.

Разработка словообразовательной разметки в НККЯ

В рамках данной темы рассматриваются методы и приемы внедрения в корпус словообразовательной разметки – одной из лингвистических разметок казахского языка. Поскольку автоматическое распознавание текста основывается на формальной структуре, показывается технология кодирования производных слов, образованных посредством синтетических и аналитических словообразовательных способов (составные слова, сложные слова, слова с аффиксами).

ABSTRACT

The monograph consists of 11 chapters dedicated to the experience of creating the subcorpora of the Kazakh language.

1. Conceptual basis of creating the National Corpus of the Kazakh language.

This chapter describes the conceptual basis for the creation of the National Corpus of the Kazakh Language, its present and future, as well as the stages of its development at the Institute of Linguistics named after. A. Baitursynuly.

2. Oral subcorpus. Since oral speech reveals and allows one to understand the linguistic traditions of a particular nation, such texts enrich the content of the national corpus. The necessity of including the following texts in the oral speech subcorpus is proved: conversational style (dialogues between family members, relatives and acquaintances, neighbors; conversations in shops, transport, public places; informal conversations on the phone), informal conversation, traditional, impromptu speech in interviews, oral presentations, pre-prepared speeches at lectures, conferences, round tables, councils, seminars and interviews, film discourse and stage performances, pre-prepared and improvised oral religious discourse, impromptu speech on television and radio programs with a known topic. Impromptu speech is true and natural. In this regard, the article describes the national characteristics of oral speech. In particular, the traditional speech of elderly people from areas with a predominantly Kazakh population is analyzed.

Oral subcorpus of the Kazakh language: creating the phonetic and prosodic marking in text corpora. In the process of research, knowledge related to the segmental and supersegmental phonetics of the Kazakh language was accumulated and systematized, and prosodic markings were prepared. Phonetic and prosodic descriptions of each native Kazakh sound are provided, and symbols of intonation components are provided. The ways of phonetic and prosodic marking of texts are considered.

Prosodic marking of texts in the oral subcorpus database has been provided, and an orthoepic norm of the speaker's speech has been created. Also, the oral subcorpus will serve as the foundation for numerous studies devoted to the phonetic-prosodic features of the word.

Phonetic-phonological and intonational transcription of the text within the oral subcorpus base.

The study examines the role of phonetic-phonological transcription in modern linguistics and the scope of its use. In modern linguistics, little attention is paid to the prosodic aspect of text transcription. The study is based on the works of academician Z.M. Bazarbaeva, who for the first time examined in detail the problem of intonation in Kazakh linguistics. In the process of text analysis, models of eight main intonemes were considered. In other words, in addition to phonetic transcription, intonation transcription of the text will also be provided in order to determine prosodic features. Also, when transcribing a text, each sentence is divided into syntagms and rhythmic groups; are marked with special signs and components such as timbre, melody, pause, tempo are described. Models of intonemes for each sentence are given.

The experience of creating the orthoepic converter.

A converter is a special program that converts files of one type into files of another type. An orthoepic converter is a program that rewrites words, phrases or texts originally written according to spelling rules in accordance with orthoepic norms.

The Institute's departments of phonetics and applied linguistics were responsible for the creation of the orthoepic converter. After analyzing existing works on phonetics, departments prepared an updated algorithm for orthoepic rules. The algorithm, consisting of 183 rules, was supported by examples.

A prepared converter can solve several important problems. Firstly, it complements the internal marking of the national corpus of the Kazakh language (<https://qazcorpus.kz/>) and correctly divides words into syllables. Secondly, he rewrites texts from the database of the oral subcorpus, guided by the norms of orthoepy. Thirdly, the converter will be released as a separate product. In addition, to achieve high-quality results when creating speech technologies, it is necessary to have automated orthoepic norms and dictionaries.

3. The onomastic subcorpus.

Institute of Linguistics named after. A. Baitursynov began developing an onomastic subcorpus, which is an independent component of the National Corpus of the Kazakh language, in the field of toponymy. Within the framework of this topic, the direction of creating an onomastic subcorpus and its principles are discussed. As a result, the main criteria for creating an onomastic subcorpus are identified. Both the national corpus of the Kazakh language and the experience of large world corpora are described. Comparisons related to the development of toponymic markings are made, and proposals are put forward. As can be seen from the experience of foreign corpora, the presence of a large number of corpora only enriches the national corpus. The emergence of the onomastic subcorpus creates many new research opportunities for onomastic scholars, as the search for texts where place names appear is accelerated. This not only simplifies the search process, but also contributes to comprehensive research in the field of onomastics. The onomastic subcorpus is, of course, necessary not only from the point of view of approval and standardization of geographical names, but also for the systematization of toponyms. The onomastic subcorpus allows preserving the national fund of toponyms and provides access to their scientific analysis, satisfying the interest of a researcher.

The technology of entering mythonyms and theonyms to the national corpus of the Kazakh language.

In the onomastic branch of Kazakh linguistics, certain studies were carried out, dissertations were defended, and dictionaries of names were compiled. However, there is no consistency in research devoted to revealing the meaning of worldviews arising from the ancient ethnic groups' knowledge of the picture of the surrounding world, which in turn is stored in names and is based on myths, mythonyms and theonyms. In this regard, within the framework of this topic, there is comparison of materials on technologies for including mythological and religious names in the corpus through carefully analyzed linguistic marking, which is publicly available on the Internet. The general views and concepts associated with the names of the corresponding mythonyms and theonyms are analyzed. Information is provided on

the technology for including mythonyms and theonyms in the national corpus of the Kazakh language. This process allows to modernize the mythological and religious space of the Kazakh language through a system of spiritual values, as well as increase the efficiency of the language teaching process.

4. From the global experience of creating historical subcorpora.

The concept of a national corpus is a tool not only for synchronic, but also for diachronic research. In this regard, the need to create a “historical subcorpus” within the national corpus of the Kazakh language is mentioned. The historical subcorpus is extremely useful and necessary when studying the language of handwritten monuments of the V-XX centuries, as well as the history, culture and literature of this period. For the first time in the field of applied linguistics, a historical subcorpus is being developed. However, there are difficulties in digitizing manuscripts that have come down to us in different graphics. The study focuses on the structure and practical application of developing a historical sub-corpus in global experience. The experience of other countries is taken into account when determining the main work that will be carried out during the development of the historical subcorpus. Issues such as identifying and collecting a database of texts from handwritten monuments of individual periods are considered; selection of obtained materials by quality and composition, grouping, processing; problems of introducing texts into the corpus and displaying them; the structure of the meta-marking of each handwritten monument. The proposed research can be used to develop a historical subcorpus of any language.

The experience of the national historical corpus of the Russian language and the demonstration of the Turkic lexical layer.

Currently, connections in the field of language and artificial intelligence, language and computer applied linguistics are rapidly developing. A continuation and new milestone of this connection is the requirement of today to consider handwritten monuments of the language as an object of computer applied linguistics. In the linguistics of developed countries, the practice of applied linguistics - corpora, including historical ones - has formed and is bearing fruit. For this purpose, the historical corpus of the Russian language was studied (applied linguistics and the construction of corpora of which are well developed) and the practice of transmitting archaic linguistic units and the manifestation of the Turkic lexical layer in it was examined. In addition, the national corpus of the Armenian language was reviewed in order to assess the experience of creating a world historical corpus. In the course of creating and improving the historical corpus of the Kazakh language, we offer the opportunity to create corpora of historical dictionaries, historical onomastics, historical terminology, historical spelling, historical grammar, historical word formation, historical phraseology, historical compatibility, historical theology, old religious texts and historical frequency. The creation of a historical corpus is one of the mechanisms to prevent any language from being absorbed by globalization. The historical corpus is a modern digital golden bridge between the library and the reader.

Lexical-grammatical marking of the historical subcorpora texts. Linguistic marking is the linguistic information provided for each unit of text in orthographic, phonetic, lexical and grammatical aspects. However, preparing the linguistic mark-

ing of texts in the historical subcorpus is an extremely complex job that requires deep research. For example, the bulk of the texts included in the historical subcorpus are written in Arabic script. These texts are transcribed into Cyrillic in different ways. The development of the historical subcorpus, in comparison with other subcorpora, is accompanied by much greater theoretical and technical difficulties. In this regard, the problem of linguistic marking of texts of the historical subcorpus was considered. In this regard, the experience of developing a historical subcorpus of the Russian language is considered; the transcription of texts written in Arabic script in different periods is analyzed; variants of transcribed words are determined; the lexico-grammatical marking of transcribed texts is clarified; the mechanisms of lexical and grammatical search for transcribed texts are described.

5. The parallel subcorpus. Basic notions of translation studies.

It is known that the origins of such a relevant, useful and important field as translation studies or the science of translation lie in the history of any nation. In the culture of every ethnos that formed a nation, translation becomes an integral part of social life - especially after the advent of writing; assumptions and opinions about the translation are expressed. Such assumptions often include a search for answers to questions about what the role of translation is in society and in cultural life, what contribution it makes to culture, and what its features are. The concept of linguistic theory of translation was formed in the twenties of the 20th century: initially this term was considered in close connection with the concept of literary translation.

«**The parallel subcorpus of the Kazakh and Russian languages: development, usage and issues**». Within the framework of this topic, a brief description of the types of parallel subcorpora is given. In a parallel subcorpus, which is being developed within the framework of the national corpus of the Kazakh language, the original text of the epic novel “The Path of Abai” by M. Auezov and its translation into Russian created by A. Kim are manually compared.

In the process of developing a parallel subcorpus, the following software tools were used – Microsoft Office Excel, Notepad++, Python, Django, MySQL.

The parallel subcorpus consists of a *database of comparable texts*, a *search system* and *information about the entered texts* (meta-marking): *author, translator, title of the work, title of the translation, time of publication of the work, time of translation, original language, target language*. The search system allows you to find the desired word by parameters such as *word, phrase, sentence, capital letter* (in Kazakh and Russian languages).

Features of translating non-verbal means in the novel «The path of Abai». The epic novel “The path of Abai” is a work of art of high cultural and linguistic value, providing extensive information about the history and life of the Kazakh people. Starting from the second half of the 20th century to the present, the novel has been constantly translated into many languages of the world. However, these translations are not made from the original, i.e. from the Kazakh language, but through the donor language, which is Russian. A controversial issue whether such a secondary translation preserves not only the original idea of the author, but also the artistic influence of the text and its linguistic and cognitive features. In this regard, the linguistic analysis of the translation of language units in the novel “Abai’s Path” is especially relevant.

Non-verbal means, which make it possible to understand the national characteristics of a people, also play an important role in any literary text from the point of view of both quantity and quality, i.e. in terms of expressive, informative meaning. The study carried out a comparative analysis of paraelements in the original text and their equivalents in the translated text. Thus, it has been revealed and supported by examples that gestures are translated in the following six ways: *units which cannot be translated (lacunae)*; *units translated verbatim*; *units translated qualitatively*; *equivalent units*; *freely translated units* and *semi-translated units*.

Lexicographic potential of the parallel text subcorpus – this topic is devoted to the study of the lexicographic potential of a parallel corpus of texts. The work pays special attention to the methods and tools of analysis, separate demonstration of original data in parallel corpora in the lexicographical aspect. The results enrich our understanding of lexicography and open up the future of contemporary research in parallel corpora and computer natural language processing.

6. Educational corpus. Issues of developing the educational corpus of the Kazakh language. Within the framework of this topic, a review of global experience in the development of educational buildings is conducted. Issues such as the types of texts covered in the English and Russian Education Corpora and the types of linguistic information provided for language learners are explored. The prerequisites for the development of the educational corpus of the Kazakh language, which is being developed at the Institute of Linguistics named after. A. Baitursynuly, are shown. Meta-marking and linguistic marking aimed at language learning and language teaching, ways and computer-software methods for adapting reference materials to the acquisition of the Kazakh language by a person learning the language are considered.

Adapting the lexical layers of the Kazakh language to the educational corpus – ways of automatization and semantization.

The National Corpus of the Kazakh language is a complex digitized version of Kazakh speech. The Kazakh language corpus is capable of providing the opportunity to learn the language, and it also consists of several subcorpora. In particular, the demand for educational corpus is increasing. The reason lies in the fact that in Kazakhstan, a multinational state, not only the state language, but also Russian and English are widely used. Therefore, representatives of other nations express a desire to find equivalents of elements of Kazakh culture in their culture and turn to translations. Also, the educational corpus will undoubtedly become an effective cognitive and linguistic base for those learning the language.

The study examines the semantization and division of the cognitive-linguistic fund introduced into the educational corpus into thematic, significant, cultural-semantic groups, its adaptation to automation corresponding to the digital system, as well as the problem of distinguishing the educational building from other subcorpora; ways of semantic classification and display of groups by meaning (adaptation to automation) are proposed.

7. Dialectological subcorpus. The technology of developing the dialectological subcorpus of the Kazakh language.

Within the framework of this topic, a review of world experience in creating dialectological corpora is conducted. In the national corpus of the Russian language,

the dialectological subcorpus is composed of materials from oral speech. The study shows the problems of phonetic transcription of these materials, their spelling, the development of prosodic markings and ways to develop automatic morphological analyzes of dialectical usages. In addition, the technology for introducing metatext marking in the Saratov dialectological corpus is described. Based on this experience in creating dialectical corpora, the development of a dialectological subcorpus of the Kazakh language is described and ways of improving it are considered. Currently, the dialectical corpus of the Kazakh language is being developed based on the material of the “Regional Dictionary of the Kazakh Language”. In this regard, the article provides a technology for processing the “Regional Dictionary” database in the process of introducing it into the subcorpus.

The meta-marking of the dialectical corpus of the Kazakh language is given at the regional, morphological and lexical levels. In particular: part of speech, region, variant, lexical meaning, grammatical transformation, phonetics and phonology. In the future, it is planned to record on audio tape and introduce into the subcorpus the materials from regional oral speech. Within the framework of this goal, proposals for organizing work are given.

8 The subcorpus of cultural-representative texts. Developing the cultural-semantic marking.

The linguistic and cultural subcorpus is considered not only as a way of in-depth acquaintance with the ethnocultural development of the Kazakh ethnic group, its spiritual and material values and the origins of ethnographic and cultural-cognitive concepts, but also as a significant tool of knowledge that describes the national unifying function of the linguistic heritage. For the first time, work is being undertaken to introduce the Linguocultural Subcorpus into the National Corpus of the Kazakh Language. The main goal of cultural-semantic marking is to determine the knowledge hidden behind linguistic units, the cultural meaning of which is to be revealed. The subcorpus is valuable because, in addition to the lexical meaning of the word, it provides its ethno-labeled meanings, ethnocultural semantics, cognitive system of knowledge and knowledge of precedent content hidden in the word.

The scientific significance of the topic is a demonstration of the prerequisites for the implementation of the Linguocultural Corpus and models for interpreting cultural codes hidden in ethno-labeled cultural units. The practical significance of the study is that it can act as a reference book on the features of the Linguocultural Subcorpus and on the algorithm for obtaining the necessary information. Techniques for disclosing cultural and linguistic information embedded in the code of cultural linguistic units covered in this subcorpus are shown.

Means of representing the ethnocultural units in the corpus.

This topic is devoted to ways of introducing into the NCKL of the descriptions of material culture in the system of ethnocultural units of the Kazakh language, as well as ways of marking and demonstrating them in the corpus.

Ethnocultural units are linguistic signs with a dominant cumulative function, which allow one to understand the national identity of the Kazakh language and preserve the language of the ethnic group, its traditions, worldview and spiritual world. In this regard, the transfer to the modern generation of the original Kazakh

picture of the world, reflected in ethnocultural units, becomes an urgent problem: an effective and accessible tool for improving the knowledge of the current generation in the era of globalization. It is known that the modern generation prefers to gain knowledge not from books, but through technological means. Thus, it is extremely important to enrich the modern generation with our national values and transfer the national fund of knowledge to the future generation.

The National Corpus of the Kazakh language is a modern innovative information tool that performs the function of preserving national cultural units and transferring them to future generations.

Studying the names of ancient titles and positions while entering the system of linguo-cultural knowledge to the national corpus.

An urgent problem is the knowledge of the worldview, consciousness, civilization and culture of a certain people through its linguistic information. The main goal of the study is to study the historically established system, hierarchy and levels of titles and positions in ancient Kazakh society, their ethnolinguistic character, their meanings, as well as the internal and external sources of their development. Therefore, of particular importance is the determination of the ethnocultural meaning of ancient names of titles and positions in the Kazakh language and the disclosure of their ethnolinguistic nature.

The lexical wealth under study not only contains cultural information about the government of the country, power, titles and positions; it stores valuable data showing the culture, mentality, worldview and knowledge, as well as the identity of the Kazakh people. Therefore, these names are a valuable heritage of a spiritual-chronical character, which bears the imprint of history and time. The names of titles and positions show the sequence of culture during the formation of the Kazakh ethnic group and describe the structure, formation and development of Kazakh society in each period of history.

Issues of annotating cultural units while developing cultural-representative subcorpora.

One of the very pressing problems of modern Kazakh society is the modernization of the spirit of the people through a comprehensive analysis of the relationship between the original culture and history, created through language and preserved in the language. The solution to this problem is directly related to the contact of the structural system of language research with the anthropocentric direction of modern Kazakh linguistics, which is actively developing today and in line with which the national picture of the world is displayed. Recent research and published lexicographic works show the relevance of using expanded information contained in culturally representative texts showing the identity of each people. Such texts, in which the unique features and authenticity of each people are realized, are the national code of the ethnos, its priceless heritage, its linguistic treasury. Therefore, a very important problem is the collection, classification, in-depth understanding and modernization of cultural texts. Thus, by collecting ancient cultural texts and creating a subcorpus of the national language, we will gain access to a modern linguistic reflection of national values. Outdated words, the meanings of which have been forgotten or faded, will be modernized and re-entered into the lexical fund.

Ethnocultural units saved in the language of the Kazakh inhabitants of China – this topic is dedicated to the analysis of ethnocultural units which can be met in the language of Kazakhs living in China, and which have not been yet included in the local dictionaries.

Semantic compatibility of the word “camel” and the ways of its automatization in the national corpus.

The study is dedicated to the analysis of the compatibility potential of the word “camel” and its semantic valency in the linguistic view. The aim of the article is to show the high compatibility of the word “camel” and to suggest the ways of its design in the subcorpus via presenting the conceptual-branch knowledge which serve as its basis. The design of the automated text that comes with the word “camel” is also carried out through the searching system. By bounding these texts to the subcorpus of “camel”, it is possible to instill the semantic marking of sentences in which this word can be met.

9. The subcorpus of proverbs and sayings: textual base, methods of demonstrating explications. Proverbs and sayings are a treasury of words, a fund of folk wisdom, formed over centuries in the national language, stored in the public consciousness and passed on from generation to generation in the form of a spiritual heritage. Since it is a philosophy formed and conceptualized through the life experiences of a people, the thoughts presented in it are characteristic of human actions and human character in general. In particular, it is undeniable that the introduction of proverbs and sayings into the consciousness of young people who do not have extensive life experience is an effective tool for creating a visionary generation. In the era of globalization, youth can be taught proverbs through advances in information technology, making proverbs and sayings in the public domain. Availability at present is the introduction of a database of proverbs and sayings into cell phones, social networks, and the development of a subcorpus of proverbs and sayings that is convenient for use. The user of the corpus can easily find the proverbs and sayings he needs through keywords or thematic groups by entering them in the search cells.

Actively used paremiologisms.

The authors of the study set the goal of identifying proverbs and sayings that are actively used at present, and revealing their modern semantics, their cultural and semantic content. As a result, it is revealed that actively used proverbs have almost completely retained their original meaning. On the other hand, proverbs and sayings that were of great importance for the life of nomadic Kazakhs (about cattle, about nomads, about weapons), but not related to modern life, are not represented among the paremiologisms of active use and gradually turn into a passive vocabulary. The authors dwell on the fact that paremiologisms are a selective combination of the ethnic character of an ethnos, its behavior, its consciousness, as well as the centuries-old experience of the people.

10. The subcorpus of advertising. The structure of advertising texts in “Kazakh” newspaper.

The main goal of the study is the classification into thematic groups and linguistic analysis of advertising texts presented in the Kazakh newspaper, organized

by the Alash party and published in 1913-1918. For this, firstly, advertising texts from 7-40 issues of the newspaper of 1913 were transcribed and graphic projections were created. The study also reviewed the history of Kazakh advertising. The sentence structure of historical advertising texts in the Kazakh language, ways of combining words, and statements are analyzed. Based on historical advertising texts published in issues 7-40 of the Kazakh newspaper in 1913, types of advertising at the beginning of the twentieth century were identified. It was determined that trade and public types of advertising were most often published in the Kazakh newspaper. The features of advertising decoration technology used at the beginning of the twentieth century are described in detail.

Peculiarities of understanding implicit advertising texts.

When understanding advertising texts, cognitive processes and mechanisms are activated in the mind of a native speaker. For this reason, we turn to the process of cognition and its mechanisms to solve linguistic problems associated with the display of advertising in the mind. Understanding an advertising text is not limited to reading the meanings clearly shown in it, but is also carried out by distinguishing hidden meanings through the process of inference. Therefore, the article also examines the process of text inference in advertising communication.

11. Developing the linguistic marking. Intangible nouns: some issues of identifying the lexical marking in the National corpus of the Kazakh language.

Intangible nouns are a lexical layer that is rarely studied in Kazakh linguistics and Turkology. Such nouns, denoting abstract concepts, properties and qualities, various stages of action and states, represent a vast layer of the Kazakh verbal fund. The main feature of intangible nouns is that they denote the speaker's attitude to the environment and to reality.

Within the framework of this topic, proposals are given for defining and delimiting the lexical-semantic marking of intangible nouns introduced into the national corpus of the Kazakh language.

Based on the classification of Akhmet Baitursynov and other famous scientists and Turkologists, intangible nouns of the Kazakh language are grouped by structure and meaning, ways of determining their lexical-semantic marking, as well as their semantic features are considered.

Ways of designing the lexical-semantic marking of nouns in the national corpus. Lexical-semantic marking of nouns is a system of signs that provides lexical-semantic information about the nouns included in the corpus. When designing this marking in the National Corpus of the Kazakh language, the "From form to meaning" setting was taken as a basis. Lexical-semantic classification of nouns: order, morphological signification – part of speech (n., adj.); word-building signification (morphemes of the word (root; root+affix; root+root); means of word-building (synthetic/analytical/lexical-semantic); form (main/motivated); content (simple/compound; hyphenated compound); morphological-semantic marking (concrete/abstract character (tangible/intangible)); animate/ inanimate character (animate/inanimate); common/proper nouns (common/proper); lexical-semantic marking: (lexical layer (native/borrowed/foreign/local/professional /historical/ archaic/new/ bookish); style pattern (colloquial/ naïve colloquial/expr/emot, curse/gratitude/ euphemism); terminologi-

cal use (aviat. anat. agricult. archaeol. art, astr. biol. bot. chem. geol. geogr. geom. hunt. jewel. lit. law, med. mil. mus. psychol. theol. sport, zool.); thematic-semantic group (macrogroup, microgroup).

Developing the lexical-semantic marking of verbs in the National corpus of the Kazakh language: global experience, classification, representation in the corpus.

The creation of a national corpus of the Kazakh language is based on the development of marking that performs automatic analysis for each level of the language. One of the complex types of marking for linguistic annotation of words is lexical-semantic marking. Within the framework of this topic, the development of lexical-semantic marking as the main type of designation in the world experience of creating corpora is considered. In particular, a review of the work of domestic and foreign researchers in the fields of computational linguistics and lexical-semantic classification is carried out, the stages of development of the lexical-semantic marking of verbs of the National Corpus of the Kazakh Language are shown, and practical foundations are explained.

Compared to the corpora of Russian, Kalmyk and other languages, the lexical-semantic marking of the National Corpus of the Kazakh language goes deeper into the meaning of the word, that is, into the sema.

Ways of implementing the lexical-semantic marking of the national corpus of the Kazakh language. This section describes work devoted to the classification of verbs into large and small semantic groups by determining their lexical-semantic, lexical-grammatical designations-codes. The question of how the automation of lexical-semantic marking of verbs is carried out is considered. The process of preparing a lexical-semantic subcorpus of verbs begins, first of all, with dividing the selected verbs into macro- and micro-groups, taking into account their semantic features. Next, ways to develop lexico-grammatical, morphological, word-formation, and connotative markings are shown.

A classification of semantic groups of verbs is created, which is necessary for the development of a lexical-semantic marking program, and ways of introducing lexical-semantic marking of verbs into the subcorpus are considered.

Issues of automatic recognition and lexicographisation of compound verbs in the corpus.

Among programs dedicated to natural language processing, programs that automatically mark up text occupy a special place. Computer programs that automate morphological analysis (taggers) in the process of work “assign” each lexical unit in the text its grammatical description. The morphological corpus analyzer does not recognize complex words as a whole unit, but considers each word separately. The study examines ways of recognizing complex verbs as integral lexical and grammatical units during automatic morphological analysis. The necessity of creating a dictionary of complex verbs is substantiated.

Developing the word-building marking in NCKL.

Within the framework of this topic, methods and techniques for introducing word-formation marking into the corpus are considered - one of the linguistic marking types of the Kazakh language. Since automatic text recognition is based on a for-

mal structure, the technology for encoding derived words formed through synthetic and analytical word-formation techniques (hyphenated compound words, compound words, words with affixes) is shown.

БІЗДІҢ АВТОРЛАР

Фазылжан Анар Мұратқызы – А.Байтұрсынұлы атындағы Тіл білімі институтының директоры, филология ғылымдарының кандидаты

Уәли Нүргелді Мақажанұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл мәдениеті бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор

Базарбаева Зейнеп Мүсілімқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор, ҰҒА академигі

Манкеева Жамал Айтқалиқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Лексикология бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор

Алдаш Айман Мырзалықызы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл мәдениеті бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор

Шойбеков Рүстембек Нұсқабекұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Этнолингвистика бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор

Малбақов Мырзаберген Малбақұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл тарихы және түркология бөлімінің бас ғылыми қызметкері, филология ғылымдарының докторы, профессор

Момынова Бағдан Қатайқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Грамматика бөлімінің меңгерушісі, филология ғылымдарының докторы, профессор

Рысберген Қыздархан Құрмашқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ономастика бөлімінің меңгерушісі, филология ғылымдарының докторы, доцент

Күдерінова Құралай Бимолдақызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ахметтану бөлімінің меңгерушісі, филология ғылымдарының докторы, профессор

Жаңабекова Айман Әбділдәқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің меңгерушісі, филология ғылымдарының докторы

Құлманов Сәрсенбай Қуантайұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың жетекшісі, филология ғылымдарының кандидаты, доцент

Ғабитханұлы Қайрат – А.Байтұрсынұлы атындағы Тіл білімі институты Лексикология бөлімінің жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Әшімбаева Нағима Мәсімаханқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Рсалиева Нұрсәуле Мақсұтқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Аманбаева Айсәуле Жантемірқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Мамырбек Гүлфар Мәжитқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ахметтану бөлімінің жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Сейтбекова Айнұр Аташбекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл тарихы және түркология бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Әмірбекова Айгүл Байдебекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Лексикология бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Әміржанова Назира Серікқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл мәдениеті бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Жұмабаева Жанар Төлендіқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Исаева Гүлсінай Сайлауқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Әлеуметтік лингвистика бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Өтебаева Эльмира Әбдіғалиқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Этнолингвистика бөлімінің меңгерушісі, филология ғылымдарының кандидаты

Солтанбекова Альфия Абдыкенқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Грамматика бөлімінің жетекші ғылыми қызметкері, филология ғылымдарының кандидаты

Бисенғали Ақмарал Зиноп-Ғабденқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың аға ғылыми қызметкері, PhD доктор

Пашан Дана Мұсабекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ономастика бөлімінің аға ғылыми қызметкері, филология ғылымдарының кандидаты

Қаршығаева Айнұр Аралбекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің аға ғылыми қызметкері, филология ғылымдарының кандидаты

Көбденова Гүлжихан Байғазықызы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл тарихы және түркология бөлімінің аға ғылыми қызметкері, филология ғылымдарының кандидаты

Үдербаев Алмас Жауынбайұлы – Қ.Сәтбаев атындағы Қазақ ұлттық техникалық университеті Қазақ және орыс тілдері кафедрасының меңгерушісі, филология ғылымдарының кандидаты, доцент

Шүленбаев Нұрлан Құныпияұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың аға ғылыми қызметкері, магистр

Садырбаева Зүбайра Боранбекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ономастика бөлімінің ғылыми қызметкері

Оспанғазиева Назгүл Бақытқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Фонетика бөлімінің ғылыми қызметкері, PhD докторант

Садық Дидар Айдарұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Ономастика бөлімінің ғылыми қызметкері, PhD докторант

Қожахметова Ақтоты Қожахметқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің ғылыми қызметкері, PhD докторант

Пиязбаева Ақмарал Бейсенқұлқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Әлеуметтік лингвистика бөлімінің ғылыми қызметкері, PhD докторант

Барменқұлова Аида Серікханқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің ғылыми қызметкері, PhD докторант

Тлегенова Гүлден Бақытқазықызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің ғылыми қызметкері, PhD докторант

Дүйсен Қуатбек Қонысбайұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Әлеуметтік лингвистика бөлімінің ғылыми қызметкері, PhD докторант

Рамазанов Талғат Бекболатұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Грамматика бөлімінің ғылыми қызметкері, PhD докторант

Пірманова Күнсұлу Қамбарбекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің кіші ғылыми қызметкері, PhD докторант

Мұрсал Айкерім Мұрсалқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің кіші ғылыми қызметкері, PhD докторант

Серікқызы Маржан – А.Байтұрсынұлы атындағы Тіл білімі институты Этнолингвистика бөлімінің кіші ғылыми қызметкері, PhD докторант

Талғатқызы Гүлнар – А.Байтұрсынұлы атындағы Тіл білімі институты Лексикология бөлімінің кіші ғылыми қызметкері, PhD докторант

Сейдамаат Әсел Қазбекқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Тіл тарихы және түркология бөлімінің кіші ғылыми қызметкері, PhD докторант

Иманғазина Меруерт Ануарқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Грамматика бөлімінің кіші ғылыми қызметкері, PhD докторант

Сейтбатқал Әрлен Берікұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Ономастика бөлімінің кіші ғылыми қызметкері, PhD докторант

Аязбаев Ғалымжан Маратұлы – А.Байтұрсынұлы атындағы Тіл білімі институтының программисі, PhD докторант

Махамбетова Айдана Мұханбайқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Әлеуметтік лингвистика бөлімінің аға лаборанты, PhD докторант

Ескендір Жарас Ертайұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Терминологиялық орталықтың аға лаборанты, магистр

Нұрлан Меруерт Ерқанатқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің аға лаборанты, гуманитарлық ғылымдар магистрі

Бейсенхан Самал Бейсенханқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімінің аға лаборанты, гуманитарлық ғылымдар магистрі

Слямбеков Қымбат Берікұлы – А.Байтұрсынұлы атындағы Тіл білімі институты Грамматика бөлімінің аға лаборанты, магистрант

Нұстаева Ардана Батырханқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Этнолингвистика бөлімінің аға лаборанты, магистрант

Кожамет Әсем Мұхитқызы – А.Байтұрсынұлы атындағы Тіл білімі институты Ахметтану бөлімінің аға лаборанты

ҚҒҰҚ QR-КОДТАРЫ



НЕГІЗГІ
КОРПУС



АУЫЗША
ІШКОРПУС



ТАРИХИ
ІШКОРПУС



ПАРАЛЛЕЛЬ
ІШКОРПУС



АХМЕТ БАЙТҰРСЫНҰЛЫ
МӘТІНДЕРІНІҢ ІШКОРПУСЫ



ДИАЛЕКТІЛІК
ІШКОРПУС



ОНОМАСТИКАЛЫҚ
ІШКОРПУС



ЖАРНАМА
ІШКОРПУСЫ



МӘДЕНИ-РЕПРЕЗЕНТАТИВТІ
МӘТІНДЕР ІШКОРПУСЫ



МАҚАЛ-МӘТЕЛДЕР
ІШКОРПУСЫ

МАЗМҰНЫ

АЛҒЫ СӨЗ (А.Фазылжан)	3
-----------------------------	---

1 ҚАЗАҚ ТІЛІ ҰЛТТЫҚ КОРПУСЫН ЖАСАУДЫҢ КОНЦЕПТУАЛДЫҚ НЕГІЗДЕРІ

1.1 Қазақ тілінің ұлттық корпусы: концептуалдық негіздері, бүгін мен болашағы (А.Фазылжан)	4
1.1.1 Қазақ тілінің ұлттық корпусының концептуалдық негіздері (А.Фазылжан) ...	4
1.1.2 ҚТҰК бірінші кезең басымдықтары туралы (А.Фазылжан)	8
1.1.3 Қазақ тілінің ұлттық корпусының бүгінгі (А.Фазылжан)	11
1.1.4 Қазақ тілінің ұлттық корпусының болашағы (А.Фазылжан)	18
1.2 Қазақ тілінің ұлттық корпусын әзірлеудің тарихи кезеңдері (А.Жаңабекова)	22

2 АУЫЗША ІШКОРПУС

2.1 Ауызша ішкорпус: сөйлеу тілінің ерекшеліктері туралы (Н.Уәли)	30
2.2 Қазақ тілінің ауызша корпусы және ауызша тіл (Қ.Күдерінова, Ә.Қожахмет) ...	40
2.3 Айтылым үдерісіне қатысты ауызша әдеби тілдің нормалары (Н.Әміржанова) ...	48
2.4 Қазақ тілінің ауызша корпусын құрудың ғылыми-практикалық мәселелері (Н.Әміржанова)	50
2.5 Қазақ тілінің ауызша ішкорпусы: мәтіндер корпусындағы фонетикалық, просодикалық белгіленімдер әзірлемесі (З.Базарбаева, А.Фазылжан, А.Аманбаева, Ж.Жұмабаева, А.Қаршығаева)	57
2.6 Ауызша ішкорпус базасындағы мәтіннің фонетика-фонологиялық және интонациялық транскрипциясы (А.Фазылжан, А.Аманбаева, Ж.Жұмабаева)	68
2.7 Орфоэпиялық конвертер жасау тәжірибесі (С. Бейсенхан, Ф. Аязбаев)	73

3 ОНОМАСТИКАЛЫҚ ІШКОРПУС

3.1 Ономастикалық корпус: антропонимдер мен топонимдердің мәдени тілдік кодын ашудың ғылыми негіздері (Н.Уәли)	80
3.2 Ономастикалық ішкорпус жасаудың әлемдік тәжірибесі (Қ.Рысберген, Н.Рсалиева, З.Садырбаева, Д.Пашан, Ә.Сейтбатқал)	95
3.3 Қазақ тілінің ономастикалық ішкорпусын әзірлеу тәжірибесі (Қ.Рысберген, З.Садырбаева, Д.Пашан, Ә.Сейтбатқал, Д.Садық)	106
3.4 Мифонимдер мен теонимдерді Қазақ тілінің ұлттық корпусына енгізу технологиясы (Қ.Рысберген, Г.Тлегенова)	115

4 ТАРИХИ ІШКОРПУС

4.1 Тарихи ішкорпус әзірлеудің әлемдік тәжірибесі (А.Сейтбекова, Ә.Сейдаман)	127
4.2 Орыс тілінің тарихи корпусын жасау тәжірибесі мен түркілік лексикалық қабаттың берілуі (Г.Көбденова, М.Малбақов)	131
4.3 Тарихи ішкорпус мәтіндерінің лексика-грамматикалық белгіленімі (А.Сейтбекова, А.Фазылжан, Ф.Аязбаев)	138

5 ПАРАЛЛЕЛЬ ІШКОРПУС

5.1 Аударматану ғылымының басты ұғым-түсініктері (А.Алдаш)	147
--	-----

5.2 Қазақ және орыс тілдерінің параллель ішкорпусы: әзірлемесі, қолданылуы және мәселелері (<i>С.Құлманов, М.Нұрлан, Ғ.Аязбаев, Н.Әшімбаева, А.Бисенғали</i>).....	153
5.3 М.Әуезовтің «Абай жолы» роман-эпопеясының орысша аудармасындағы бейвербалды амалдар (<i>М.Иманғазина</i>).....	162
5.4 Параллель ішкорпусқа алынған аударма мәтіндерді талдау (<i>Н.Рсалиева</i>).....	171
5.5 Параллель мәтіндер корпусының лексикографиялық әлеуеті (<i>К.Пірманова</i>).....	176

6 ОҚУ КОРПУСЫ

6.1 Қазақ тілінің оқу корпусын әзірлеу мәселелері (<i>А.Жаңабекова, Г.Тлегенова</i>).....	182
6.2 Қазақ тілінің лексикалық қабатын оқу корпусына бейімдеу: семантизация және автоматтандыру жолдары (<i>А.Әмірбекова</i>).....	189

7 ДИАЛЕКТІЛІК МӘТІНДЕР ІШКОРПУСЫ

7.1 Қазақ диалектологиясы бойынша зерттеулердегі негізгі бағыттар (<i>А.Үдербаев</i>)	197
7.2 Диалектілік мәтіндер ішкорпусы: әзірлемесі мен технологиясы (<i>А.Жаңабекова, А.Қожахметова, Г.Тлегенова</i>).....	201

8 МӘДЕНИ РЕПРЕЗЕНТАТИВТІ МӘТІНДЕР ІШКОРПУСЫ

8.1 Этнолексика – ұлт болмысының тілдік дерегі (<i>Ж. Манкеева</i>).....	212
8.2 Ұлттық корпуста ұлттық мәдениеттің этномәдени кодтарын ашу мәселелері (<i>Н.Уәли</i>)	218
8.3 Лингвомәдени ішкорпустың мәтіндік-лексикографиялық базасын әзірлеу (<i>М.Серікқызы</i>)	230
8.4 Мәдени-семантикалық белгіленім әзірлемесі (<i>Ә.Өтебаева, Г.Исаева, М.Серікқызы</i>).....	235
8.5 Мәдени-репрезентативті ішкорпусты әзірлеудегі мәдени бірліктерді аннотациялау проблемалары (<i>Ж.Манкеева, Н.Шүленбаев, А.Бисенғали</i>)	243
8.6 Қазақ лексикасындағы жаңғырған атаулардың мәдени-семантикалық сипаты (<i>Р.Шойбеков</i>)	250
8.7 Этномәдени бірліктердің корпуста берілу жолдары (<i>Ж.Манкеева, А.Барменқұлова, А.Жаңабекова</i>).....	256
8.8 Лингвомәдени білімдер жүйесін ұлттық корпусқа енгізуде көне атауларды зерттеу (<i>Н. Әшімбаева, А.Жаңабекова</i>)	263
8.9 Диаспора қазақтары тіліндегі этномәдени бірліктер (<i>Қ.Ғабитханұлы</i>)	274
8.10 «Түйе» сөзінің семантикалық тіркесімділігі және ұлттық корпуста автоматтандыру тәсілдері (<i>А. Әмірбекова, Г. Талғатқызы</i>)	279

9 МАҚАЛ-МӘТЕЛДЕР ІШКОРПУСЫ

9.1 Мақал-мәтелдер ішкорпусы: мәтіндік базасы, түсіндірме беру әдістері (<i>А.Жаңабекова</i>)	289
9.2 Белсенді қолданыстағы паремиологизмдер (<i>Ә.Өтебаева, М.Серікқызы, А.Нұстаева</i>).....	292

10 ЖАРНАМА ІШКОРПУСЫ

10.1 «Қазақ» газетіндегі жарнама мәтіндерінің құрылымы (<i>Қ.Күдерінова, Г.Мамырбек</i>)	297
--	-----

10.2 Жарнама мәтіндер ішкорпусын әлеуметтік лингвистика аспектісінде зерттеу (<i>Г.Исаева, Ж.Ескендір</i>)	307
10.3 Қазақ тіліндегі жарнама мәтіндерінің тілдік ерекшеліктері (<i>Ә.Сейдаман, Қ.Дүйсен, А.Махамбетова</i>)	316
10.4 Жарнама корпусы: сауда жарнамасындағы мәтіндердің жалпы сипаттамалары (<i>А.Қаршығайева, Н.Оспанғазиева</i>)	322
10.5 Жарнамада берілген имплицитті мәтіндерді түсіну ерекшелігі (<i>А.Пиязбаева</i>) ..	328

11 ЛИНГВИСТИКАЛЫҚ БЕЛГІЛЕНІМ ӘЗІРЛЕМЕЛЕРІ

11.1 Дерексіз зат есімдер: Қазақ тілінің ұлттық корпусында лексикалық белгіленімді анықтаудың кейбір мәселелері (<i>А.Алдаш</i>)	333
11.2 Зат есімдердің лексика-семантикалық белгіленімін ұлттық корпуста рәсімдеу жолдары (<i>А.Әмірбекова</i>)	342
11.3 Қазақ тілінің ұлттық корпусындағы етістіктердің лексика-семантикалық белгіленім әзірлемесі: әлемдік тәжірибе, классификациялау, корпуста белгілеу (<i>Б.Момынова, М.Имангазина, Қ.Слямбеков</i>)	350
11.4 Қазақ тілінің ұлттық корпусында етістіктің лексика-семантикалық белгіленімін қою жолдары (<i>А.Солтанбекова, Т.Рамазанов</i>)	362
11.5 Корпуста күрделі етістіктерді автоматты тану мәселелері (<i>А.Жаңабекова, А.Мурсал</i>)	372
11.6 ҚТҰК-дағы сөзжасамдық белгіленім әзірлемесі (<i>А.Жаңабекова, К.Пірманова</i>) ..	377
Пайдаланылған әдебиеттер	383
Глоссарий	397
Қысқаша аңдатпа	404
Краткая аннотация	415
Abstract	427
Біздің авторлар	438
ҚТҰК QR-кодтары	442

Ұжымдық ғылыми монография

**ҚАЗАҚ ТІЛІНІҢ
ҰЛТТЫҚ КОРПУСЫН ӨЗІРЛЕУ
ТӘЖІРИБЕСІ
(1-кезең)**

Беттеген: Т. Есжанов
Мұқабасын көркемдеген: Б. Альдибаев

Басуға 24.10.2023 жылы қол қойылды. Пішімі 70x90 1/16
Қаріп түрі «Times New Roman». Қағазы офсеттік.
Көлемі 28 б.т. Таралымы 150 дана.

Редакцияның мекенжайы:
050010, Алматы қ. Құрманғазы көшесі, 29
ҚР ҒЖЖМ ҒК А.Байтұрсынұлы атындағы Тіл білімі институты
Тел./факс: 8 (727) 272-55-47, 272-69-61