

[https://doi.org/10.53360/3080-3861-2026-1\(5\)-19](https://doi.org/10.53360/3080-3861-2026-1(5)-19)  
FTAMP: 17.01.45

**К.К. Муканова, Ә.М. Қуанышева\***

Шәкәрім университеті,

071412, Қазақстан Республикасы, Семей қ., Глинка к-сі, 20 А

\*ORCID: 0009-0000-0177-4997

\*e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com)

## **CORPUS-BASED LEARNING: ФИЛОЛОГИЯЛЫҚ БІЛІМ БЕРУДІҢ ЖАҢА МОДЕЛІ**

**Аңдатпа.** Қазіргі заманғы әлемдік үрдісте лексикалық дерекқор толық цифрландырылған жағдайда, тілді меңгеру әлдеқайда тиімді болады. Қазақ тілінің ұлттық корпусы – қазақ сөзінің цифрлық нұсқасы. Күн сайын сұраныс артып келе жатқан корпус тіл үйренушілер мен зерттеушілер үшін сенімді және тиімді лингвистикалық базаға айналды. Мақалада оқу корпусындағы лингвистикалық жүйені цифрлық форматқа бейімдеу, корпусның құрылымымен танысу, іздеу және талдау тапсырмаларын орындау, жалпылау және сипаттау әдістері қарастырылады. Корпустық лингвистиканың ғылыми-әдістемелік негіздері, қағидалары мен зерттеу тәсілдері олардың тілдік деректерді жүйелі талдауы ұсынылады. Сонымен қатар филологиялық білім беру үдерісінде корпустық әдістерді қолданудың инновациялық қырлары, теориялық және қолданбалы зерттеулерде жанашылдыққа қалай ықпал еткенін көрсетеді. Зерттеу барысында қазақ тілінің лингвистикалық корпустарындағы грамматикалық белгілердің салыстырмалы талдауы жасалады.

Білім алушылардың корпустармен тиімді жұмыс істеуі үшін корпус сауаттылығын қалыптастырудың маңызы ерекше екені көрсетіледі. Осы мақсатта зерттеу корпустың құрылымымен танысу, іздеу және талдау тапсырмаларын орындау, шағын мәтіндік деректер негізінде ішкі корпустар құру, сондай-ақ корпустық деректерді пайдаланудың негізгі қағидаларын меңгеру кезеңдерін қамтиды. Талдауға негіз болған тілдік материалдар «Қазақ тілінің ұлттық корпусының кіші корпустар» базасынан алынды.

Қорытынды бөлімде корпустық зерттеулердің болашақ бағыттары мен оқу үрдісінде дамытуға қатысты ғылыми тұжырымдар мен әдістемелік ұсынымдар беріле отырып, бұл бағыттағы зерттеулерді әрі қарай тереңдету қажеттілігі атап өтілген.

**Тірек сөздер:** корпустық лингвистика, corpus-based learning, филологиялық білім беру, цифрлық гуманитаристика, коммуникативтік құзыреттілік, инновациялық модель.

### **Кіріспе**

Қазіргі жаһандану және цифрлық трансформация жағдайында білім беру жүйесіндегі құрылымдық өзгерістер филологиялық пәндерді оқытудың мазмұны мен әдістемесін ғылыми негізде жаңартуды талап етеді. Нақты деректерге сүйене отырып тілдік құбылыстарды талдау, материалдарды электронды форматқа көшіру, цифрландыру арқылы зерттеу жұмыстарын жеңілдету және тиімділігін арттыру, халықаралық тіл білімінің даму үрдісіне ілесу, бәсекеге қабілетті болу сияқты талаптар бүгінгі күні қазақ тіл білімінде компьютерлік бағдарламалар арқылы жұмыс істейтін мәтіндік және сөздік дерекқорларды қалыптастыруда маңыздылығын көрсетеді.

Тілдік деректерді жүйелі жинақтау, өңдеу және талдау қағидаларына негізделген корпустық лингвистика тіл білімі ғылымының жас әрі қарқынды дамып келе жатқан салаларының бірі. Корпустық лингвистика әдістемесін түсіну үшін «корпус» ұғымын нақтылау қажет. Корпус – лингвистикалық зерттеулер жүргізу мақсатында пайдаланылатын, түпнұсқа мәтіндер мен тілдік қолданыс үлгілерінен құралған компьютерлендірілген дерекқор. Оқу корпусын дамытуға және оны арнайы зерттеуге деген қызығушылықтың себебі қоғамның сұранысынан туындайды. Қазіргі қазақ қоғамында мемлекеттік тілді үйренушілер ғана емес, сонымен қатар қазақ тілі ортасында да дұрыс жазуға, сөйлеуге және сөйлемдер құруға, әрбір сөзді мағынасына сәйкес қолдануға күш салынууда. Осы тұрғыдан алғанда, жалпы көпшілік қазақ тілінің сауатты қолданылуы мен дамуына жанашырлықпен қарайтынын байқадық. Себебі, қоғамдық сұраныс бойынша сөздердің мағынасын түсіндіру, жаңа сөздердің аудармасы, әдеби тіл мен жергілікті немесе ауызекі тілдердің қолданылу ортасы мен мәртебесі, қатесіз жазу мен сөйлеу нормалары, сөздердің антонимі, синонимі,

омонимдік сипаты, бейнелі қолданылуы және фразеологиялық қордың басқа тілдегі мағынасы және т.б. басымдыққа ие болып келеді.

Корпустық лингвистика – тіл деректерін ірі көлемде зерттеу, яғни транскрипцияланған сөйлемдер немесе жазба мәтіндердің өте кең жинақтарын компьютерлік талдау арқылы зерттеу [1]. Корпустық әдістер тілдік материалды эмпирикалық деректер негізінде зерттеп, дәстүрлі лингвистикалық талдауды жаңа технологиялық тәсілдермен толықтырады.

Әлемдік лингвистикада тілдік зерттеулер жүргізуде әртүрлі ұлттық ірі мәтіндік корпустары қалыптастырылған. Мысалы, «Ағылшын тілінің Британ ұлттық корпусы (BNC, 100 млн.) (<http://www.natcorp.ox.ac.uk/>), америкалық ұлттық корпус (ANC, 15 млн.) (<https://anc.org>), түрік тілінің ұлттық корпусы (TNC, 50 млн.) (<https://www.tnc.org.tr>), орыс тілінің ұлттық корпусы (RNC, 2 млрд.) (<http://www.ruscorpora.ru>) және т.б. корпустар – тіл ресурстарын бір платформа аясына жинай алған ірі құрылымдар» [2].

Заманауи технологиялардың дамуы корпустық талдауды оқу үрдісіне енгізуді жеңілдетіп, оқытушы мен білім алушы арасындағы өзара байланыстың жаңа формасын қалыптастырды. Дерекке негізделген оқыту тәсілдері студенттердің зерттеушілік қабілеті мен сыни ойлау дағдыларын дамытып, цифрлық ортада деректермен жұмыс істеу қабілеті қалыптасады. Мақалада жоғары оқу орындарында филология мамандығы студенттерінің тілдік құзыреттерін дамытуда корпус лингвистикасы құралдарын қолданудың әдістемелік негіздері қарастырылады.

Зерттеудің мақсаты – филологиялық пәндерді оқытуда корпустық талдау әдістерінің тиімділігін анықтау және студенттердің тілдік заңдылықтарды нақты деректер негізінде зерттеу қабілетін арттыру. Зерттеу барысында корпустық лингвистиканың филологиялық білім берудегі теориялық негіздері талданып, оқыту процесінде қолданылатын корпустық әдістер жүйеленіп, тиімділігі тәжірибелік тұрғыдан қарастырылады.

#### Әдеби шолу

Шетелдік ғылыми әдебиеттерде корпустық лингвистика көбінесе мәтіндік корпустарды талдау негізінде тілдік зерттеулер жүргізудің әдіснамасы ретінде түсіндіріледі. Корпустық лингвистиканың негізін қалаушы британ ғалымы J. Sinclair корпустық деректерге сүйене отырып тілдік заңдылықтарды анықтаудың теориялық негіздерін ұсынып, corpus-driven әдісін дамытты [3]. Американдық лингвисті В. Douglas корпустық талдаудың грамматикалық және функционалдық зерттеулерде қолданудың тиімділігін дәлелдеді [4]. Дерекке негізделген оқыту тұжырымдамасы білім алушылардың тілдік заңдылықтарды өз бетінше анықтау қабілетіне арналған.

Орыс тіл білімінің ғалымы В.А. Плуноянның зерттеуінде, «корпус – тіл зерттеудің қуатты құралы ғана емес, сонымен қатар зерттеушіні мәтінді теориялық рефлексияның басты нысаны ретінде қарастыруға бағыттайтын жаңа идеология» [5].

Қазақ тіл біліміндегі компьютерлік лингвистиканың зерттелуі А. Жұбановтың 2002 ж. қорғаған «Основные принципы формализации содержания казахского текста» атты докторлық монографиясынан басталады. Ғалым: «Әлемдік корпустық лингвистиканың даму сипаты – ұлттық толық мәтіндерді арнайы зерттеу нысаны етіп алуды қажет етуде. Қазақ мәтіндерінің стильдік, құрылымдық, мағыналық, функционалдық 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», – деп жазған [6]. Кейін бұл бағытты А. Жаңабекова, Б. Карбозова, С. Құлманов, А. Фазылжан т.б. ғалымдар корпус деректерін ғылыми тұрғыда пайдалану мәселелерін зерттеді.

Отандық ғылымда қазақ тілінің ұлттық корпустары QazCorpus ([qazcorpus.kz](http://qazcorpus.kz)), QazCorpora ([qazcorpora.kz](http://qazcorpora.kz)), Қазақ тілі корпусы 2 ([issai.nu.edu.kz](http://issai.nu.edu.kz)) және Алматы қазақ тілі корпусы ([web-corpora.net](http://web-corpora.net)) сияқты жобалар қазақ тілінің табиғи мәтіндік қорын жүйелеуге, тілдік деректерді сандық форматта талдайды. (Қоңыратбаева, 2022). Бұл корпустық жобалар қазақ тілінің табиғи мәтіндік қорын жүйелеуге, тілдік деректерді сандық форматта талдауға және лингвистикалық зерттеулердің заманауи түрін дамытуға бағытталған. 70 миллион

сөзқолданысты қамтитын А. Байтұрсынұлы атындағы Тіл білімі институты базасындағы Қазақ тілінің ұлттық корпусы бірнеше ішкорпустардан тұрады. (<https://qazcorpus.kz>) Мұндай құрылым ұлттық тілдің әртүрлі функционалдық стильдерін кешенді түрде зерттейді. Публицистикалық мәтіндерді қамтитын шамамен 24 миллионға жуық сөзқолданыс жинақталған Ш. Шаяхметов атындағы «Тіл-Қазына» ұлттық ғылыми-практикалық орталығының базасындағы корпустың мақсаты – табиғи тілдік ресурстарды жинақтау, тілдік деректерді жүйелеу және пайдаланушыларға қолжетімді цифрлық платформа ұсыну (<https://new.qazcorpora.kz/landing-page>). Қазақ тілінің ауызша тілдік деректерін жинақтауға бағытталған жобалардың бірі – Назарбаев университеті *Ақылды жүйелер мен жасанды интеллект институты* (ISSAI). Корпустың базасында 2000-нан астам қатысушыдан жиналған шамамен 300 сағаттық ауызша тілдік материал қамтылған. Аталған ресурс қазақ тіліндегі сөйлеуді тану және синтездеу технологияларын дамытуға бағытталған лингвистикалық база (<https://issai.nu.edu.kz/kk/home-kaz/>). Алматы қазақ тілі корпусы шамамен 40 миллион сөзқолданысты қамтып, оның құрамына көркем әдебиет, ғылыми және публицистикалық мәтіндер енгізілген. ([http://web-corpora.net/KazakhCorpus/search/?interface\\_language=kz](http://web-corpora.net/KazakhCorpus/search/?interface_language=kz)).

Аталған ресурстарды білім беру үрдісінде пайдалану филологиялық пәндерді оқытуда дерекке негізделген әдістерді енгізуге, студенттердің тілдік талдау және академиялық жазу дағдыларын дамытуға жағдай жасайды. Осыған байланысты корпустық лингвистика құралдарын оқыту жүйесіне енгізу ғылыми-әдістемелік тұрғыдан өзекті мәселе ретінде қарастырылады.

### **Материалдар мен әдістер**

Корпустық лингвистика филологиялық пәндерді оқытуда білім алушылардың теориялық білімін тәжірибемен ұштастырып, тілдік құбылыстарды дерек негізінде түсіндіріп, ғылыми талдау дағдыларын дамытады. Корпус құралдарымен жұмыс істеу арқылы студенттер тілдік материалды дербес талдап, белсенді тіл үйренушіге айналып, үздіксіз ізденіс дағдыларын қалыптастыра алады.

Ғалым А. Жұбанов былай түсіндіреді: «Корпустық лингвистиканың екіжақтылық сипатын (корпусты құру және оны пайдалану) оның нысанының екіжақтылығымен түсіндіруге болады. Себебі, біріншіден, нысан деп отырғанымыз, корпустық лингвистика үшін де және басқа да тілдік пәндер үшін де – сөйлеу материалы, ал, екіншіден, корпустық лингвистика қызметінің нәтижесі» [7]. Мұндай тәжірибе корпусқа сүйене отырып жаңа тілдік заңдылықтарды ашуға және өз теориялық тұжырымдарын қалыптастырады.

Зерттеу материалы ретінде студенттердің тілдік қателіктерін тексеретін, академиялық мәтін үлгілерін және әртүрлі жанрадағы мәтіндерді қамтитын электрондық корпустар пайдаланылды. Бұл деректер тілдік бірліктердің қолданыс жиілігін, мағыналық ерекшеліктерін және контекстік қызметін анықтауға негіз болды. Қолданыстағы оқу бағдарламалары мен оқу-әдістемелік материалдар талданды. Зерттеуде сипаттамалық, салыстырмалы және корпустық талдау әдістері қолданылды. Корпус негізінде сөздердің жиілік көрсеткіштері, коллокациялық байланыстары мен конкорданс қатарлары айқындалды. Конкорданстар – іздеу сөзі немесе сөз тіркесі негізінде жасалған мәтін үзінділері Конкорданс талдауы белгілі бір сөз формасының бастапқы контекстінде қолданылуының көптеген мысалдарын бір жерде жинауға мүмкіндік береді. Коллокациялық талдау немесе синтаксистік және семантикалық тұрғыдан үйлесімді бірліктің сипаттамаларын бөлісетін екі немесе одан да көп сөздердің тіркестері корпуста жиі бірге қолданылатын сөздерді анықтау үшін қолданылады [8].

Зерттеуде жиі кездесетін грамматикалық және лексикалық қателер жүйеленіп, корпустық талдау нәтижесінде алынған деректерді филологиялық пәндердің (грамматика, лексикология, мәтінтану, стилистика, тіл мәдениеті) мазмұнын жаңаша ұйымдастыруда пайдалану ұсынылды. Әр тұжырым лингвистикалық әдістерге сүйене отырып жасалып, мысалдармен дәлелденді.

Корпустық лингвистиканы қолданудың әдістемелік негіздері аясында корпустарды оқу үрдісіне синхронды және диахронды тәсілдер арқылы кешенді енгізу айрықша мәнге ие.

Оқытушы үшін корпусқа негізделген жаттығулар әзірлеу қажетті тілдік бірліктерді іріктеуге, оқу мақсатына сай мысалдар қорын қалыптастыруға, тапсырмаларды дайындау барысында жиі кездесетін қателіктердің алдын алуға және материалдың жүйелілігін сақтауға көмектеседі. Корпусты пайдалану барысында белгілі бір тілдік деңгейге сәйкестендіру, мақсатты аудиторияның ерекшеліктерін ескеру және тиімді әдіс-тәсілдерді қолдану назарға алынды.

### Нәтижелер мен талқылау

Корпус құралдарын оқу үрдісіне енгізу бірнеше оң нәтижелерге алып келді. Білім алушылар тілдік үлгілерді, сөз тіркестерін және грамматикалық құрылымдарды терең түсініп, тілдің дәлірек қолданылуына қол жеткізеді.

Корпус лингвистикасының әдіснамасы мен әдістерінің дамуының арқасында тілдерді сипаттамалық зерттеу зерттеушілер мен кез келген басқа пайдаланушы үшін қолжетімді, оны корпустар кез келген тілдік деңгейлерде: фонетикалық-фонологиялық, грамматикалық, семантикалық, прагматикалық және басқа да деңгейде қолдайды [9]. Әртүрлі корпустарды қолдану аймақтық тіл ерекшеліктерін қосуды қамтамасыз етіп, тіл үйретуде мәдени сананы және сезімталдықты арттырады. Қазақ тілінің оқу корпусы екі мақсатты жүзеге асыруы керек. Біріншісі - қазақ тілді азаматтарға тілдік сауаттылықты одан әрі жақсарту, яғни күмән тудыратын дұрыс сөзді табу, тілдік ережелерді қарастыру үшін оңтайлы және қолжетімді мүмкіндік беретін іздеу құрылғысын әзірлеу. Екіншісі - қазақ тілін үйренушілерге арналған электрондық қосымша. Оған тілдің деңгейі мен тақырыбына байланысты сөздік, тілашар, бейне, аудиовизуалды материалдар, мәтіндер (тақырыптар) кіреді [10].

Корпустық лингвистика құралдарын қазақ тілі мен әдебиеті мамандығы студенттерін оқыту процесіне енгізу олардың тілдік құзыреттерін арттыруға, грамматикалық және лексикалық қателерді жүйелеуге, сондай-ақ оқыту әдістемесін жетілдіреді. Практикалық жұмыстар барысында конкорданс жолдарын талдау, сөз жиілігін анықтау және тұрақты тіркестерді салыстыру әдістері студенттердің тілдік заңдылықтарды өздігінен эмпирикалық бақылау арқылы түсінуіне ықпал етті. Сонымен қатар, корпустық талдау мәдени сананы дамытуға және аймақтық тіл ерекшеліктерін ескеруге жағдай жасайды.

Алайда, корпус құралдарын қолдануда бастапқыда үйрену қиындықтары және қажетті оқыту ресурстарының жетіспеушілігі сияқты проблемалар байқалды. Осы мәселелерді жан-жақты оқыту және қолдау арқылы шешу корпустық талдау әдістерінің тиімділігін одан әрі арттыра алады және филологиялық пәндерді оқытуда инновациялық тәсілдердің дамуына үлес қосады.

Корпустық құралдарды игерудегі қиындықтарға қарамастан, Қазақ тілі ұлттық корпусының мультимедиялық мүмкіндіктері тілдік материалды функционалды-контекстік тұрғыдан талдауда жоғары тиімділік көрсетеді. Мұның айқын дәлелі ретінде ауызша ішкорпус (oral corpus) құрамындағы бейнематериалдарды қолдануды атап өтуге болады (1-сурет).



- Сәлеметсіз бе?
- Осы сіз қайда жұмыс істейсіз? жиі көремін.
- Мен емханада, дәрігермін. Ал сіз ше?
- Оқытушымын. Студенттерге қазақ тілінен сабақ берем.
- Оқытушы болған қиын емес пе?
- Жо-жоқ, өзіме ұнайт. Ал дәрігер болған ше?
- Болады, қиындықтары да бар.
- Иә, рас... Дәрігер болу жауапты жұмыс. Менің анам да дәрігер.
- Танысқаным өте қуаныштымын.
- Мен де, сау болыңыз!
- Көріскенше.

1-сурет. Қазақ тілі ұлттық корпусының ауызша ішкорпусындағы сұхбат үлгісі мен оның транскрипцияланған мәтіні

Берілген мысалда екі кейіпкер арасындағы мамандық пен жұмыс орны туралы диалог бейнеленген. Бұл үзіндіні филологиялық тұрғыдан талдау барысында бірнеше заңдылықтарды анықтадық:

• **Коммуникативтік стратегиялар:** сөйлеушілердің «Осы сіз қайда жұмыс істейсіз?», «Оқытушы болған қиын емес пе?» деген сұрақтары арқылы тілдік ортадағы этикеттік нормалар мен сұрақ қою манерасының табиғилығы айқын көрінеді.

• **Лексика-семантикалық қолданыс:** диалогта кездесетін «дәрігер», «емхана», «оқытушы», «жауапты жұмыс» сияқты лексемалардың мәтін ішіндегі байланысы студенттерге мамандықтарға қатысты сөздік қорды контекст арқылы меңгеруге болады.

• **Прагматикалық аспект:** сөйлеушінің «Иә, рас...» деп келісуі немесе «Жо-жоқ, өзіме ұнайды» деп жауап беруі ауызша сөйлеу тіліндегі интонациялық реңк пен эмоциялық бояуды көрсетеді. Бұл жазбаша мәтіндегі статикалық мысалдарға қарағанда, тілдің жанды динамикасын түсінуге көмектеседі.

Зерттеу барысында анықталғанындай, ауызша ішкорпустың бейнематериалдары филологиялық пәндерді оқытуда визуалды және аутентивті контекст қалыптастырады. Мысалдағы транскрипция мен видеоның қатар берілуі студенттердің когнитивтік қабылдауын жеңілдетеді. Бұл ретте корпус тек мәліметтер базасы ретінде емес, тілдік бірліктердің прагматикасын зерттейтін «тірі зертхана» рөлін атқарады.

Дегенмен, мұндай инновациялық тәсілдерді енгізу үшін оқытушылардың цифрлық құзыреттілігін арттыру және корпустық менеджментті жетілдіру қажеттілігі туындайды. Бұл бағыттағы жұмыстар филология саласындағы дәстүрлі лингвистикалық талдауды заманауи дата-талдау (data-driven learning) деңгейіне көтеруге жол ашады.

Корпустық лингвистика грамматика мен сөздік қорды оқытуда да маңызды рөл атқарады. Дәстүрлі грамматиканы оқыту көбіне нақты тілдік қолданысты толық көрсете бермейтін абстрактілі ережелерге сүйенеді. Ал корпустық деректерді енгізу арқылы мұғалімдер білім алушыларға грамматикалық құрылымдардың контекст ішіндегі шынайы мысалдарын ұсына алады. Бұл білім алушыларға грамматиканың икемділігі мен вариативтілігін түсінуге көмектесіп қана қоймай, жазбаша және ауызша тілдің айырмашылықтарын айқындауға көмектеседі. Сол сияқты, корпусқа негізделген сөздік қорды оқыту тәсілі студенттерді сөздердің табиғи қолданыс аясымен таныстырады, яғни олардың басқа сөздермен тіркесу арқылы мағыналы бірліктер құрайтынын көрсетеді. Шынайы тілдік деректермен жұмыс істеу білім алушылардың сөз қолданысын тереңірек түсінуіне және жалпы тілдік еркіндігін арттыруына ықпал етеді.

Филологиялық пәндерді оқытуда корпустық лингвистиканың мүмкіндіктері кең. Лексикология курсына корпустар сөздердің жиілік көрсеткішін, мағыналық реңктерін, қолданыс аясын анықтауға мүмкіндік береді. Мысалы, белгілі бір сөздің әртүрлі стильдегі мәтіндерде қолданылуын салыстыру арқылы оның функционалдық сипаты ашылады.

Морфология мен синтаксисті оқытуда корпус материалдары грамматикалық формалардың қолданыс ерекшелігін көрсетуге тиімді. Студенттер нақты мысалдар негізінде тілдік нормалар мен ауытқуларды талдай алады.

Стилистика және мәтін лингвистикасы сабақтарында корпустар көркем, ғылыми, публицистикалық мәтіндердің тілдік ерекшеліктерін салыстырмалы түрде зерттеуге мүмкіндік береді. Бұл тәсіл теориялық білімді практикалық дағдымен ұштастырады.

Корпустық лингвистикада жазбаша немесе ауызша мәтіндердің дәстүрлі мәтіндік талдау түрлерінен өзіндік ерекшеліктері бар. Корпустық зерттеулерде мәтін тұтастай қарастырылмайды, керісінше зерттеу мақсатына сәйкес оның жекелеген үзінділері немесе сөз, сөз тіркесі, сөйлем сияқты әртүрлі деңгейдегі тілдік бірліктер талданады. Мұнда мәтін мазмұнын түсіндіру емес, қайталанатын тілдік құбылыстар мен тілдік қолданыс үлгілерін анықтау басты назарда болады. Корпустық әдістер тілді жүйе ретінде қарастырып, тілдік талдауда сандық және сапалық тәсілдерді қатар қолдануға негізделеді. Соның нәтижесінде сөздердің немесе сөз тіркестерінің мағынасы олардың тілдік контекстінде қарастырылып, тілдік қолданыс ерекшеліктері арасындағы байланыс та айқындалады.

Әдістерге талдау, салыстыру, үлгілеу, контент-анализ және студенттердің өздік зерттеу әрекеттерін ұйымдастыру кіреді. Мысал ретінде студенттің жазған сөйлемін қарастырайық: *«Ғылыми мақала жазғанда ойларды дұрыс орналастыру қажет»*. Корпус талдауы көрсеткендей, академиялық мәтіндерде жиі қолданылатын құрылым: *«Ғылыми мақалада ойларды логикалық және жүйелі түрде орналастыру маңызды»*. Мұндағы айырмашылық: студенттің нұсқасында жалпы «дұрыс орналастыру» деген тіркес қолданылған, ал корпус үлгісі нақты **логикалық және жүйелі түрде** деген анықтамалық маркерлер арқылы академиялық стильге жақындатады. Осылайша, студент корпусының көмегімен **дұрыс сөз тіркестерін таңдап**, мәтін құрылымын академиялық стандартқа сәйкестендіре алады. Бұл әдіс грамматика мен лексикаға ғана емес, **коммуникативтік дағдылар мен жазу мәдениетіне** де оң әсерін тигізеді.

Корпусқа негізделген әдістемелік тәсіл студенттердің өздігінен үйренуін ынталандырады. Студенттер корпусар арқылы нақты мәтіндермен жұмыс істей отырып, тілдік құрылымдарды, сөз тіркестерін және дискурстық маркерлерді анықтап, қолдануды үйренеді. Бұл өз кезегінде оқу процесін студенттің қызығушылығына сай бейімдеуге, сабақтарды нақты тілдік деректерге сүйене отырып құруға және жазу дағдыларын жетілдіреді. Жазу дағдыларын дамыту – корпус лингвистикасының студенттердің тілдік қабілеттерін жетілдіруде маңызды саласы. Академиялық және кәсіби жазу корпусын талдау арқылы студенттер мәтінді құрылымдауда қолданылатын байланыстырғыш құралдар, ауысу сөздері және дискурстық маркерлерді байқай алады. Бұл тілдік үлгілерді түсіну оларға жазбаларын ұғынықты әрі жүйелі етуге көмектеседі. Сонымен қатар, корпус талдауы әртүрлі жанрлардың дәстүрлерін анықтауға көмектеседі, бұл студенттерді әртүрлі академиялық пәндерге сәйкес жазу стиліне бейімдейді. Бұл корпус негізінде тілдік теорияны тексеру мен қолданудың нақты мысалы. Қате талдау – корпус лингвистикасын қолданудың тағы бір тиімді тәсілі. Студенттердің жазбаларының корпусын жинап, мұғалімдер жиі кездесетін қателерді жүйелі түрде анықтап, жеке қажеттіліктерге сәйкес кері байланыс беруге және әлсіз тұстарды жетілдіруге арналған оқу материалдарын жасай алады. Мысалы, егер корпус талдауы студенттердің шылауды жиі қате қолданатынын көрсетсе, мұғалімдер дұрыс шылауды қолдануға арналған жаттығуларды дайындай алады. Мұндай тәсіл корпус негізінде бұрынғы тілдік теорияны тексерудің әрі студенттің нақты тілдік дағдыларын жетілдірудің айқын мысалы.

Корпус лингвистикасы тілдік бағалау құралдарын жетілдірсе, дәстүрлі бағалау көбінесе оқушылардың тілдік қабілеттерін толық көрсете алмайтын жеке тапсырмаларға сүйенеді. Ал корпусқа негізделген бағалау шынайы тілдік қолдану мен нақты өмірлік қарым-қатынасқа негізделіп, оқушылардың тілдік дағдыларын жан-жақты қарастырады. Корпус деректері арқылы жаңа бағалау әдістерін жасап, тілдік дағдыларды нақты әрі кешенді түрде зерттейді. Корпус анализінің нәтижелері бойынша оқытушы жиі кездесетін қателіктерді анықтап, бағытталған кері байланыс бере алады. Тілдік үлгілерді үздіксіз бақылау оқу материалдарын жаңартып отыруға және студенттердің академиялық мәтін жазу сапасын арттыруға көмектеседі. Бұл тәсіл филологиялық пәндерді оқытуда теориялық білім мен практикалық қолдануды байланыстыра отырып, студенттердің тілдік құзыреттілігін, сыни ойлау қабілетін және мәтіндік құрылымдарды дұрыс қолдану дағдыларын дамытуды мақсат етеді. Зерттеу барысында филологиялық пәндерді оқытуда корпусық лингвистиканы қолдану негізгі әдістемелік бағыт ретінде қарастырылды: оқу материалдарын құрастыру, грамматикалық және лексикалық ерекшеліктерді көрсету, сондай-ақ студенттердің коммуникативтік дағдыларын дамыту.

### **Қорытынды**

Оқытушы үшін корпусық лингвистика оқу материалдарын дәлелді деректерге сүйене отырып құруға мүмкіндік беріп, оқыту процесін интерактивті әрі деректерге негізделген сипатқа ие етеді. Білім алушыларға бұл тәсіл академиялық және кәсіби жазу дағдыларын жетілдіруге, тілдік қателіктерді түзетуге және мәтіндерді контекстке сай құрастыруға жағдай жасайды. Қорытындылай келе, корпусық лингвистика филологиялық пәндерді оқытудың әдістемелік базасын нығайтып, инновациялық және тиімді білім беру орталарын

қалыптастыруға үлес қосады. Болашақта оның рөлі одан әрі кеңейіп, тіл білімін сапалы түрде жетілдіру және студенттердің тілдік құзыреттілігін арттырудағы шешуші құралға айналатыны сөзсіз. Корпустық лингвистика – филологиялық пәндерді оқыту мен зерттеудің болашағы зор заманауи әдісі. Студенттерге нақты тілдік материалмен жұмыс істеуге, аналитикалық ойлауды дамытуға, тілдік деректерді дәлелмен түсіндіруге көмектесетін заманауи құрал. Бұл – әсіресе сандық лингвистика, корпус лингвистикасы, қолданбалы лингвистика салаларында орны ерекше.

Зерттеу нәтижелері көрсеткендей, «Corpus-based learning» (корпусқа негізделген оқыту) моделі филологиялық білім беруді трансформациялаудың пәрменді құралы. Тілді статикалық ережелер жиынтығы ретінде емес, динамикалық қолданыстағы құбылыс ретінде оқыту студенттерге нақты тілдік үлгілермен тікелей жұмыс істеуге мүмкіндік береді. Бұл тәсіл грамматикалық, лексикалық және синтаксистік құрылымдарды контексте игеруге, сондай-ақ мәтіндік үйлесімділік пен дискурстық маркерлерді функционалды тұрғыдан тиімді пайдалануға жағдай жасайды.

Талдау барысында анықталған негізгі нәтижелерді жинақтай келе, бірнеше маңызды тұжырым жасауға болады. Біріншіден, корпус құралдарын пайдалану студенттердің тек теориялық лингвистикалық білімін тереңдетіп қана қоймай, тілдік фактілерді өз бетінше саралауға, оларды сыни тұрғыдан бағалауға және шынайы коммуникацияда еркін қолдануға мүмкіндік беретін кешенді коммуникативтік құзыреттілігін қалыптастырады. Екіншіден, корпустық технологияларды оқу үдерісіне жүйелі түрде енгізу дәстүрлі оқыту әдістерін заманауи электрондық білім беру ресурстарымен байытып, филологиялық пәндерді оқытудың сапалы әрі инновациялық жаңа моделін құруға негіз қалайды. Үшіншіден, бұл технологиялық шешімдерді педагогтарды даярлау жүйесіне интеграциялау болашақ филолог-мұғалімдердің цифрлық дағдыларын арттырып, олардың кәсіби даярлығын заманауи жаһандық талаптар деңгейінде қамтамасыз етудің басты кепілі. Сөз соңында айтарымыз «Corpus-based learning» тәсілі тілдік білім берудің қолданбалы сипатын күшейтіп, филология саласындағы мамандарды даярлаудың тиімді стратегиясына айналатыны сөзсіз.

### Әдебиеттер

1. Mcenery T., Hardie A. Corpus linguistics. Method, theory and practice. – Cambridge University Press. 2012, – 292 p.
2. Қоңыратбаева Ж.М. Қазақ тілінің академиялық корпусын әзірлеу: мақсаты, міндеті, маңызы (гуманитарлық ғылымдар бағытында) // Қарағанды университетінің хабаршысы. Филология сериясы. – 2025, 30, 3(119), – Б. 88-98. <https://doi.org/10.31489/2025Ph3/88-98>
3. Sinclair J. (Ed.) Corpus, Concordance, Collocation. Oxford, UK: Oxford University Press. 1991.
4. Douglas B. Co-occurrence Patterns among Collocations: A Tool for Corpus-Based Lexical Knowledge Acquisition. – 1993. – 19(3), – P. 531-538.
5. Плунгян В.А. Корпус как инструмент и как идеология: О некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. – 2008. – № 2 (16). – С. 7-20.
6. Жұбанов А. Қазақ әдеби тілінің электрондық корпусын түзудің теориялық бастаулары // Тіл және мәдениет: тілдің антропоэзеттік парадигмасы. Профессор Ж.А.Манкееваның 60 жылдығына арналған республикалық ғылыми-теориялық конференцияның материалдары. – Алматы: А.Байтұрсынұлы атындағы Тіл білімі институты, 2010. – Б. 191-197.
7. Жұбанов А.Қ. Корпустық лингвистика / А.Қ. Жұбанов, А.Ә. Жаңабекова. – Алматы: Қазақ тілі, 2017. – 336 б.
8. Kuchumova G., Muhamejanova D., Ibrayeva A. Corpus linguistics as a methodology for conducting empirical language studies and its development in Kazakhstan // Bulletin of the Karaganda University. Philology Series. – 2024. – Vol. 29 No. 2 (114), – P. 96-105. <https://doi.org/10.31489/2024ph2/96-105>.

9. Pirmanova K.K., Turekhanova A., Ashimova N. Theory and methodology of the world's national linguistic corpora // Bulletin of the Kazakh State University of Languages and Literatures named after Abylai Khan: Philological Sciences Series. – 2023. – Vol. 69 No. 2. – P. 158-181. <https://doi.org/10.48371/PHILS.2023.69.2.011>.
10. Amirbekova A.B., Konyrova A.T., Kaiyrbekova U.S. Methods of automating texts in the national corpus of the Kazakh language. // Bulletin of the Kazakh State University of Languages and Literatures named after Abylai Khan: Philological Sciences Series. – 2023. – Vol. 70 No. 3. – P. 11-24. <https://doi.org/10.48371/PHILS.2023.70.3.001>.
11. Қазақ тілінің ұлттық корпусы. – [Электрондық ресурс]. – Қолжетімділігі: <https://qazcorpus.kz>
12. Қазақ тілі ұлттық корпусы. – [Электрондық ресурс]. – Қолжетімділігі: <https://qazcorpora.kz/>
13. Қазақ тілі корпусы. – [Электрондық ресурс]. – Қолжетімділігі: <http://test-test-issai.nu.edu.kz/kz-speech-corpus/>
14. Алматы қазақ тілі корпусы. – [Электрондық ресурс]. – Қолжетімділігі: [http://webcorpora.net/KazakhCorpus/search/?interface\\_language=ru](http://webcorpora.net/KazakhCorpus/search/?interface_language=ru)

### References

1. McEnery T., Hardie A. Corpus Linguistics. Method, Theory and Practice [Corpus Linguistics. Method, Theory and Practice]. – Cambridge: Cambridge University Press, 2012. – 292 p.
2. Qongyrtbaeva Zh.M. Qazaq tilinin akademiyaalyq korpusyn ázirleu: maqstsy, mindeti, manyzy (gumanitarlyq gylymdar bagytynda) [Development of the Academic Corpus of the Kazakh Language: Purpose, Tasks, and Significance (in Humanities)] // Qaragandy universitetinin habarshysy. Filologiya seriyasy. – 2025. – Vol. 30, No. 3 (119). – P. 88-98. <https://doi.org/10.31489/2025Ph3/88-98> (in Kazakh)
3. Sinclair J. (Ed.) Corpus, Concordance, Collocation. – Oxford: Oxford University Press, 1991. 207 p.
4. Douglas B. Co-occurrence Patterns among Collocations: A Tool for Corpus-Based Lexical Knowledge Acquisition. – 1993. – Vol. 19 (3). – P. 531-538.
5. Plungyan V.A. Korpus kak instrument i kak ideologiya: O nekotorykh urokakh sovremennoi korpusnoi lingvistiki [Corpus as a Tool and as an Ideology: On Some Lessons of Modern Corpus Linguistics] // Russkii yazyk v nauchnom osveshchenii. – 2008. – No. 2 (16). – P. 7-20. (in Russian)
6. Zhubanov A. Qazaq ádebiet tilinin elektrondyq korpusyn tüzudin teoriyalyq bastaulary [Theoretical Foundations of Developing the Electronic Corpus of the Kazakh Literary Language] // Til zhane madeniet: tildin antropoózehtik paradigmasy. – Almaty: A. Baitursynuly atyndagy Til bilimi instituty, 2010. – P. 191-197. (in Kazakh)
7. Zhubanov A.Q., Zhanabekova A.A. Korpus tyq lingvistika [Corpus Linguistics]. – Almaty: Qazaq tili, 2017. – 336 p. (in Kazakh)
8. Kuchumova G., Muhamejanova D., Ibrayeva A. Corpus Linguistics as a Methodology for Conducting Empirical Language Studies and Its Development in Kazakhstan // Bulletin of the Karaganda University. Philology Series. – 2024. – Vol. 29, No. 2 (114).
9. Pirmanova K.K., Turekhanova A., Ashimova N. Theory and Methodology of the World's National Linguistic Corpora // Bulletin of the Kazakh State University of Languages and Literatures named after Abylai Khan: Philological Sciences Series. – 2023. – Vol. 69, No. 2. – P. 158-181. <https://doi.org/10.48371/PHILS.2023.69.2.011>.
10. Amirbekova A.B., Konyrova A.T., Kaiyrbekova U.S. Methods of Automating Texts in the National Corpus of the Kazakh Language // Bulletin of the Kazakh State University of Languages and Literatures named after Abylai Khan: Philological Sciences Series. – 2023. – Vol. 70, No. 3. P. 11-24. <https://doi.org/10.48371/PHILS.2023.70.3.001>.
11. Qazaq tilinin ulttyq korpysy [National Corpus of the Kazakh Language]. – Electronic resource. – Available at: <https://qazcorpus.kz> (accessed 01.2026). (in Kazakh)

12. Qazaq tili ulttyq korpysy [Kazakh Language National Corpus]. – Electronic resource. – Available at: <https://qazcorpora.kz/> (accessed 01.2026). (in Kazakh)
13. Qazaq tili korpysy [Kazakh Language Corpus]. – Electronic resource. – Available at: <http://test-test-issai.nu.edu.kz/kz-speech-corpus/> (accessed 01.2026). (in Kazakh)
14. Almaty qazaq tili korpysy [Almaty Kazakh Language Corpus]. – Electronic resource. – Available at: [http://webcorpora.net/KazakhCorpus/search/?interface\\_language=ru](http://webcorpora.net/KazakhCorpus/search/?interface_language=ru) (accessed 01.2026). (in Kazakh)

**К.К. Муканова, Ә.М. Қуанышева\***

Шәкәрім университет,  
071412, Республика Казахстан, г. Семей, ул. Глинки, 20 А

\*ORCID: 0009-0000-0177-4997

\*e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com)

## **CORPUS-BASED LEARNING: НОВАЯ МОДЕЛЬ ФИЛОЛОГИЧЕСКОГО ОБРАЗОВАНИЯ**

**Аннотация.** В современных мировых тенденциях, когда лексические базы данных полностью цифровизированы, процесс изучения языка становится значительно более эффективным. Национальный корпус казахского языка представляет собой цифровую версию казахского слова. Корпус, спрос на который ежедневно возрастает, стал надежной и эффективной лингвистической базой для изучающих язык и исследователей. В статье рассматриваются вопросы адаптации лингвистической системы учебного корпуса к цифровому формату, ознакомления со структурой корпуса, выполнения поисковых и аналитических заданий, а также применения методов обобщения и описания. Представлены научно-методологические основы корпусной лингвистики, её принципы и исследовательские подходы, направленные на системный анализ языковых данных.

Кроме того, показаны инновационные аспекты применения корпусных методов в процессе филологического образования и их влияние на развитие теоретических и прикладных исследований. В ходе исследования проводится сравнительный анализ грамматических показателей в лингвистических корпусах казахского языка.

Отмечается особая значимость формирования корпусной грамотности для эффективной работы обучающихся с корпусами. С этой целью исследование включает этапы ознакомления со структурой корпуса, выполнения поисковых и аналитических заданий, создания внутренних корпусов на основе небольших текстовых данных, а также освоения основных принципов использования корпусных данных. Языковой материал для анализа был взят из базы «малых корпусов Национального корпуса казахского языка».

В заключительной части представлены научные выводы и методические рекомендации по дальнейшему развитию корпусных исследований в учебном процессе, а также подчеркивается необходимость углубления исследований в данном направлении.

**Ключевые слова:** корпусная лингвистика, corpus-based learning, филологическое образование, цифровая гуманитаристика, коммуникативная компетенция, инновационная модель.

**К.К. Mukanova, A.M. Kuanysheva\***

Shakarim University,  
071412, Republic of Kazakhstan, Semey, Glinka Str. 20 A

\*ORCID: 0009-0000-0177-4997

\*e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com)

## **CORPUS-BASED LEARNING: A NEW MODEL OF PHILOLOGICAL EDUCATION**

**Abstract.** In the context of contemporary global trends, when lexical databases are fully digitized, language learning becomes significantly more effective. The National Corpus of the Kazakh Language represents a digital version of Kazakh vocabulary. As the demand for corpus resources continues to grow, it has become a reliable and efficient linguistic database for both language learners and researchers. This article examines the adaptation of the linguistic system of an educational corpus to a digital format, familiarization with the corpus structure, the implementation of search and analytical tasks, and the application of generalization and descriptive methods. The scientific and methodological foundations of corpus linguistics, as well as its principles and research approaches for the systematic analysis of linguistic data, are discussed.

In addition, the article highlights innovative aspects of applying corpus-based methods in philological education and demonstrates how they contribute to advancements in theoretical and applied research. The study also presents a comparative analysis of grammatical features found in linguistic corpora of the Kazakh language.

The importance of developing corpus literacy for students to work effectively with corpus resources is emphasized. For this purpose, the research includes stages such as familiarization with corpus structure, performing search and analytical tasks, creating internal corpora based on small textual datasets, and mastering the main principles of using corpus data. The linguistic material used for the analysis was taken from the “sub-corpora” database of the National Corpus of the Kazakh Language.

The conclusion presents scientific findings and methodological recommendations for further development of corpus studies in the educational process and emphasizes the need to deepen research in this field.

**Keywords:** corpus linguistics, corpus-based learning, philological education, digital humanities, communicative competence, innovative model.

#### **Авторлар туралы мәлімет**

**Муканова Каршыга Кайроллаевна** – педагогика ғылымдарының кандидаты, Шәкәрім университеті қазақ филологиясы кафедрасының доценті, Қазақстан Республикасы, Семей, e-mail: [mukanova.k@mail.ru](mailto:mukanova.k@mail.ru), ORCID: <https://orcid.org/0000-0002-8812-7625>.

**Қуанышева Әйгерім Мұратқызы\*** – Шәкәрім университеті «Қазақ тілі мен әдебиеті мұғалімдерін даярлау» ББ магистранты, Қазақстан Республикасы, Семей, e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com), ORCID: <https://orcid.org/0009-0000-0177-4997>.

#### **Сведения об авторах**

**Муканова Каршыга Кайроллаевна** – кандидат педагогических наук, доцент кафедры казахской филологии Шәкәрім Университет, Республика Казахстан, Семей, e-mail: [mukanova.k@mail.ru](mailto:mukanova.k@mail.ru), ORCID: <https://orcid.org/0000-0002-8812-7625>.

**Қуанышева Әйгерім Мұратқызы\*** – магистрант Шәкәрім Университет по образовательной программе «Подготовка учителей казахского языка и литературы», Республика Казахстан, Семей, e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com), ORCID: <https://orcid.org/0009-0000-0177-4997>.

#### **Information about authors**

**Karshyga Mukanova** – candidate of pedagogical sciences, associate professor department of kazakh philology Shakarim University, Republic of Kazakhstan, Semey, e-mail: [mukanova.k@mail.ru](mailto:mukanova.k@mail.ru), ORCID: <https://orcid.org/0000-0002-8812-7625>.

**Aigerim Kuanysheva\*** – master’s student at Shakarim University, Educational Program «Training of kazakh language and literature teachers», Republic of Kazakhstan, Semey, e-mail: [aikerim.k.m.03@gmail.com](mailto:aikerim.k.m.03@gmail.com), ORCID: <https://orcid.org/0009-0000-0177-4997>.

*Редакцияға енуі 14.01.2026*

*Жариялауға қабылданды 27.03.2026*

[https://doi.org/10.53360/3080-3861-2026-1\(5\)-20](https://doi.org/10.53360/3080-3861-2026-1(5)-20)  
IRSTI:16.21.37

#### **D.Y. Yermurat**

Intellectual development center Zerdeli Education,  
071410, Republic of Kazakhstan, Semey, 67/1 Shakarim street  
ORCID: 0009-0001-7297-0300  
e-mail: [danaermuratova@gmail.com](mailto:danaermuratova@gmail.com)

### **PRONUNCIATION CORRECTION METHODS FOR LEARNERS OF ENGLISH AS A FOREIGN LANGUAGE: A SYSTEMATIC REVIEW**

**Abstract.** This article presents a systematic review of pronunciation correction methods for learners of English as a Foreign Language (EFL). The aim of the study is to compare traditional and technology-based