

Г.Ә. Сәрсекес¹, Б.Е. Акмагамбетова², Ж.Қ. Сентірбек¹

¹Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, 010000, Қазақстан Республикасы

²М. Қозыбаев атындағы Солтүстік Қазақстан университеті, Петропавл, 150000, Қазақстан Республикасы

АКАДЕМИЯЛЫҚ КОРПУС ЖӘНЕ ОНЫ ҚАЗАҚ ТІЛІНДЕ ӘЗІРЛЕУДІҢ ӨЗЕКТІЛІГІ МЕН МАҢЫЗДЫЛЫҒЫ

Мақалада академиялық корпустар туралы және академиялық қазақ тілінің жазбааша корпусын әзірлеу барысындағы ғылыми зерттеулер нәтижелері баяндалады. Ғылыми зерттеудің мақсаты – академиялық қазақ тілінің жазбааша корпусын әзірлеудің өзектілігі мен маңыздылығын негіздеу. Мақалада қазіргі қазақ корпус лингвистикасының жайы, біздің ғылыми зерттеуіміздің аталған салаға қосар үлесі, академиялық корпустар туралы, академиялық корпустар құрудың шетелдік тәжірибесі, сондай-ақ қазақ тілінде академиялық корпус құрудың өзектілігі мен маңыздылығы талқыға түседі.

Бұл жұмыста академиялық жазбааша қазақ тілінің корпусын әзірлеу аясында жүргізілетін ғылыми зерттеулердің нәтижесінде тақырыптың негізгі ұғымдары анықталады және олардың қазіргі қазақ тілді ғылыми зерттеулердегі қолданыстары ашылады. Академиялық корпустар құру ісінің шетелдік тәжірибесіне ағылшын тіліндегі академиялық корпустарға типі, мазмұны, құрылымы, көлемі және репрезентативтілігі тұрғысынан сараптама жасалды.

Зерттеудің нәтижелерін қазақ тілінің корпус лингвистикасы және компьютер лингвистикасы салаларында, сонымен қатар аталған пәндерді жоғары оқу орындарында оқытуда теориялық және практикалық материалдар ретінде пайдалануға болады. Академиялық жазбааша қазақ тілінің корпусын әзірлеу және ұсыну ұлттық тілдің ғылыми әлеуеті мен қолданбалы мүмкіндіктерін арттыруға, қазақ тілінің халықаралық кеңістікте кеңінен насихатталуына, сондай-ақ гуманитарлық ғылымдарды цифрландыру ісін жеделдетуге үлес қосады. Академиялық корпус және бұл тақырыпта жүргізілген ғылыми зерттеулер академиялық қазақ тілі ұғымын бекемдеуге, академиялық лексиканы айқындауға, академиялық мәтінді жан-жақты зерттеуге мүмкіндік береді. Қазақ тілін академиялық мақсатта оқыту ісіне үлес қосады.

Кілт сөздер: академиялық, академиялық корпус, академиялық мәтін, қазақ тілі, жазбааша корпус, академиялық мақсатта оқыту, цифрландыру, гуманитарлық ғылымдар

НЕГІЗГІ ЕРЕЖЕЛЕР

Академиялық корпус лингвистикалық корпустардың бір түрі болып табылады. Корпустар электронды мәтіндердің жиынтығын ғана емес, түрлі лингвистикалық құралдарды және қосымшаларды қоса қамтитын үлкен ресурс ретінде қолданбалы лингвистикада және лингводидактикада, сондай-ақ басқа да салаларда корпуста негізделген ғылыми зерттеулер жүргізуде үлкен қолданбалы маңызға ие болып отыр. Академиялық корпустар академиялық салалардың әртүрлі пәндеріндегі тілдің қолданысы, нақтылай айтқанда академиялық жазбааша мәтіндерде тілдің қалай қолданылатыны туралы түсінік береді. Академиялық корпустар академиялық мәтіндердің түрлі жанрлары мен регистрлерінен жиналған үлкен көлемі арқылы академиялық жазбааша және сөйлеу тілінің табиғи қалпын тануға көмектеседі. Корпус деректері арқылы академиялық мәтіндердің басқа мәтін түрлерінен ерекшеленетін өзіндік белгілері мен сипаттамаларын алуға болады.

Бұл зерттеуде академиялық жазбаша қазақ тілінің корпусын әзірлеу жобасы аясында жүргізілген ғылыми зерттеулердің нәтижелері баяндалады. Бұл тақырыптағы ғылыми зерттеулерде *академиялық* атрибут ұғымы негізгі кілт сөз болып табылады. *Академиялық* сөзінің тірек сыңар болуымен бірқатар тақырыптық негізгі фразалар жасалады. *Академиялық мәтін, академиялық дискурс, академиялық жазу/академиялық жазылым, академиялық корпус, академиялық қазақ тілі* т.т. осылардың қатарынан. Зерттеу мақаламызда осындай терминдік тіркестердің қазіргі қазақ тілді ғылымда қаншалықты және қалай қолданылып жүргеніне назар аудартылады. Қазақ тілінің ғылым тілі ретінде мәртебесін нығайтуда және оның академиялық салалардағы қолданысын кеңейтуде академиялық қазақ тілінің корпусы маңызды рөл атқарады. Сонымен қатар мұндай корпус қазақ тілін екінші тіл ретінде оқыту әдістемесінде де маңызды рөлге ие.

Академиялық жазбаша қазақ тілінің корпусын әзірлеу ісі осы тақырыпта ғылыми зерттеулер жүргізумен қатар жүреді. Академиялық корпус құру тақырыбында ғылыми зерттеулер жүргізу барысында академиялық корпустар құрудың халықаралық тәжірибесі мен жалпы корпус құрудың отандық тәжірибесі қарастырылады және салыстырылады.

Тұжырымдай айтқанда, мақалада көтерілетін өзекті мәселе – қазақ тілінде академиялық корпус құрудың маңыздылығы мен қажеттілігін, оны әзірлеу және көпшіліктің пайдалануына ұсынудың қолданбалы лингвистикаға, лингводидактикаға және гуманитарлық ғылымдарды цифрландыруға айтарлықтай үлес қосатынын негіздеу болды.

КІРІСПЕ

Қазіргі жаһандану заманында ақпараттық технологиялардың қарқынды дамуымен түрлі саяси-әлеуметтік, мәдени және экономикалық салалардағы ақпарат ағыны ерекше жылдамдықпен таралуда. Қоғамдағы ақпарат ағымының бұлайша тез таралуы тіл арқылы жүзеге асатындықтан, тілдің қызметі кеңейіп, функционалдық мүмкіндіктері арта түсуде. Технологиялардың қарыштап дамуымен *табиғи тілді өңдеу* ұғымы қалыптасып, қазірде қолданбалы лингвистикада пәнаралық зерттеу түрі ретінде үлкен қызығушылыққа ие болып отыр. Табиғи тілді өңдеудің жарқын мысалы лингвистикалық корпустар құру болып табылады.

Корпустар құру ісі халықаралық тәжірибеде тіл ғылымы мен компьютерлік ғылымдарда орасан зор нәтижелер беріп отыр. Корпус лингвистикасы – бұл адамзат тілін өңдеуде немесе табиғи тілдің қолданысын дәл көрсетуде компьютерлерді қолдануға бағытталған пәнаралық сала. Корпус лингвистикасы машинада оқылатын мәтіндердің үлкен жиынтығы арқылы тілді зерттеуге орасан мүмкіндіктер ашады. Машинада оқылатын мәтіндер *корпус* (латынның *corpus*, көпше *corpora* сөзінен алынған) деп аталады. Корпус – лингвистикалық зерттеулер үшін құрылған «электронды түрдегі тілдік мәтіндер фрагменттерінің жиынтығы» [1, 19]. Корпус лингвистикасы соңғы онжылдықта табиғи тілді компьютерлерде өңдеудің керемет мүмкіндіктерінің арқасында жақсы дамыды. Компьютерлердің және машинада оқылатын мәтіндердің болуы деректерді тез және оңай алуға, сондай-ақ ол деректерді талдау үшін қолайлы форматта ұсынуға мүмкіндік берді.

Компьютерлік технологиялардың мүмкіндіктерін пайдалана отырып, қазіргі қазақ тілінің академиялық жазбаша корпусын құру – аса маңызды әрі қажетті іс. Отандық тәжірибеде академиялық корпус(тар) жоқ. Біз әзірлейтін академиялық қазақ тілінің жазбаша корпусы осы олқылықтың орнын толтырады. Қазақ тілінің корпустарын жасауда алғашқы ұлттық ізденістердің барын атап өтуіміз қажет. Л.Н. Гумилев атындағы Еуразия ұлттық университеті, әл-Фараби атындағы Қазақ ұлттық университеті және А. Байтұрсынұлы атындағы Тіл білімі институты әзірлеген корпустар бар. ҚазҰУ ғалымдары құрған *Алматы қазақ тілі корпусы* мен А. Байтұрсынұлы атындағы Тіл білімі институты әзірлеген *Қазақ тілінің ұлттық корпусы* пайдалануда ашық қолжетімді. Ағылшын тілінің корпустарымен салыстыра қарасақ, қазақ тілінің қолда бар корпустары саусақпен санарлықтай аз және олардағы сөз саны әлі де болса аз. Бұл сандық көрсеткіш қазақ тіліндегі корпустарды әлі де

жетілдіре түсу қажеттілігін көрсетеді. Осы орайда корпустардың санын арттыру ғана емес, олардың түрлілігін арттыру да маңызды. Академиялық жазбаша қазақ тілінің корпусы сол түрлілікке үлес қосады.

Бұл зерттеуде, алдымен, академиялық қазақ тілінің жазбаша корпусын әзірлеу аясында жүргізілген ғылыми зерттеулердің негізінде бірқатар салалық негізгі ұғымдар анықталады. Google Scholar платформасында зерттеу жүргізу арқылы негізгі ұғымдарды белгілейтін сөздер мен фразалардың қолданысына талдау беріледі. Онан кейін академиялық корпустар құрудың халықаралық тәжірибесі сарапқа салынады. Соңында, қазақ тілінде академиялық корпус әзірлеудің өзектілігі мен маңыздылығы негізделіп, дәлелденеді.

МАТЕРИАЛДАР МЕН ӘДІСТЕР

Бұл зерттеудің негізгі материалдары қолжетімді шетелдік академиялық корпустар және ол корпустарды әзірлеу бойынша жарық көрген ғылыми зерттеу мақалалары, кітап тараулары, вебсайт материалдары, сондай-ақ корпус әзірлеуші ғылыми жоба тобымен сұхбаттасу деректері болды. Материалдарды іріктеуде ағылшын тіліндегі корпустардың ішінен зерттеу мақсатына сай тек академиялық корпус түрлері таңдалды және олармен танысу үшін олардың ашық қолжетімді болуы ескерілді. Дегенмен кейбір корпустардың ашық қолжетімді болмауына немесе жартылай қолжетімді болуына байланысты ол корпустардың қалай әзірленгені туралы белгілі бір ғылыми жұмыс тобының жариялаған ғылыми зерттеулері пайдаланылды. Сонымен қатар академиялық корпусты әзірлеп, пайдалануға берген ғылыми жоба тобымен бетпе-бет жүздесіп, сұхбаттасудың нәтижесінде алынған материалдар да зерттеу мақаламызда пайдаланылды. Айталық, осы тақырыпта ғылыми зерттеу жұмыстарын жүргізу барысында Британдық академиялық жазбаша ағылшын тілінің корпусын [2] әзірлеген Ұлыбританияның Ковентри университетінің ғылыми жұмыс тобы мүшелерімен кездесіп, сұхбаттасу мүмкін болды.

Зерттеу мәселесі – қазақ тілінде академиялық корпус әзірлеу ісінің маңыздылығын негіздеу. Қазақ тілінде қазіргі күнде академиялық корпус жоқ, сондықтан корпустың бұл түрін әзірлеп, көпшіліктің пайдалануына ұсыну өзекті әрі маңызды. Зерттеу тақырыбының жаңашылдығы да сонда. Бұл корпус қазіргі қолданыстағы қазақ тілді корпустар ресурстарындағы бос тұрған орынды толтырады.

Зерттеу кезеңдері мына сатылардан тұрды. Алдымен, академиялық корпустар әзірлеудің халықаралық тәжірибесі сарапқа салынды. Шет тілдерінде жасалған академиялық корпустармен жұмыс жасап, оларды жобалау және құру тәжірибесін бөліскен корпус әзірлеуші топтың ғылыми зерттеулерімен танысып шықтық. Академиялық ағылшын тілінің корпусын әзірлеген Ковентри университетіне барып, корпус әзірлеген ғылыми жоба тобымен кездестік. Британдық жазбаша ағылшын тілінің академиялық корпусын әзірлеген ғалымдардың іс-тәжірибелері мен кеңестері біздің корпус құрудағы ісімізге бағыт берді, кейбір сұрақтарымыздың жауабын алуға көмектесті. Келесі кезеңде, отандық тәжірибе сарапқа салынды. Алматы қазақ тілі корпусы мен Қазақ тілінің ұлттық корпусындағы деректер қорына сараптама жасадық. Академиялық корпустың нақты академиялық мәтіндерден тұру сипаты бұл корпусты аталған екі корпустан ерекшелейтін тұсы екеніне көз жеткіздік. Тек қана академиялық мәтіндерді қамтитын корпус сол мәтіндердің табиғи белгілері мен ерекшеліктерін анықтауға мүмкіндік беретіні сөзсіз. Осы кезеңде қазақ тілінің академиялық корпусына кіретін мәтіндерді іріктеу, тақырып аясындағы негізгі ұғымдарды анықтау және сипаттау, сондай-ақ ағылшын тіліндегі академиялық корпустар әзірлеудің халықаралық тәжірибесін зерттеу жүзеге асты. Ағылшын тілінен құрылған академиялық корпустарға мазмұны, құрылымы, репрезентативтілігі және көлемі сияқты сипаттамалары бойынша сараптама жасалды.

НӘТИЖЕЛЕР

Академиялық жазбаша қазақ тілінің корпусын әзірлеу жобасы аясында жүргізілген ғылыми зерттеулердің нәтижесінде академиялық корпус құру ісіне қатысты ұғымдар анықталды және олардың қазіргі қазақ тілді ғылымдағы қолданылу жиілігіне зерттеу жасалды. Соның нәтижесінде *академиялық* сөзінің қолданылу жиілігі *ғылыми* сөзінің қолданылу жиілігіне қарағанда әлдеқайда төмен екені көрінді. *Академиялық* сөзі – бұл зерттеудің кілт сөзі. Бұл сөз академиялық корпусар тақырыбындағы терминдерге тірек сыңар болып, салалық ұғымдарды жасауға негіз болады. Зерттеу барысында сондай ұғымдардан *академиялық корпус*, *академиялық қазақ тілі*, *академиялық мәтін*, *академиялық дискурс*, *қазақ тілін академиялық мақсатта оқыту* тіркестерінің қолданысына зерттеу жүргізілді. Нәтижесінде *академиялық корпус* ұғымының қазіргі қолданыста жоқ екенін, *академиялық қазақ тілі* және *қазақ тілін академиялық мақсатта оқыту* ұғымдарының өте сирек жұмсалатынын анықтадық. Мұндай негізгі ұғымдардың қолданылу жиілігіне зерттеу біздің қолға алып отырған жоба тақырыбымыздың жаңашылдығын, қазақ тілінде академиялық корпус құру ісінің маңыздылығын және қажеттілігін дәлелдей түсті.

Отандық тәжірибеде академиялық корпусар құру ісі бүгінге дейін қолға алынбағандықтан, салыстырмалы түрде академиялық корпусардың халықаралық тәжірибесі сарапталды. Сараптама әлемдік тәжірибеде академиялық корпусардың ағылшын тілінен өте көп құрылғанын көрсетті. Мұның себебі де түсінікті, қазіргі күнде ағылшын тілі халықаралық ғылым мен білім беру процесінде үлкен басымдыққа ие. Сонымен қатар ағылшын тілін ана тілі және шет тілі ретінде пайдаланушылардың саны әлем бойынша өте жоғары екені белгілі. Ағылшын тілінен әзірленген академиялық жазбаша корпусарға шолу жасалды. Шолудың мақсаты оқырманға академиялық корпусар типі, құрылымы мен құрамы, көлемі, лингвистикалық тұрғыдан жетілдірілуі жайлы ақпарат беру болды. Шетелдік академиялық корпусар құру ісі қазақ тілінің академиялық корпусын жобалау мен әзірлеу ісінде бағыт-бағдар беретін тәжірибе алаңы болатыны сөзсіз.

Отандық академиялық корпусы әзірлеу және ұсыну бірқатар мәселелердің шешім табуына ықпал етуімен маңызды. Олардың бастылары: гуманитарлық ғылымдарды цифрландыруға, сондай-ақ қазақ тіл білімінде қолданбалы лингвистика саласын дамытуға үлес қосады; академиялық мақсатта қазақ тілін оқыту ісін жандандыра түседі; қазақ тілін шет тілі ретінде меңгерту және оқыту ісіне, яғни лингводидактикаға үлес қосады; академиялық корпусқа негізделген ғылыми зерттеулерді жүргізуді арттырады және түрлендіреді.

ТАЛҚЫЛАУ

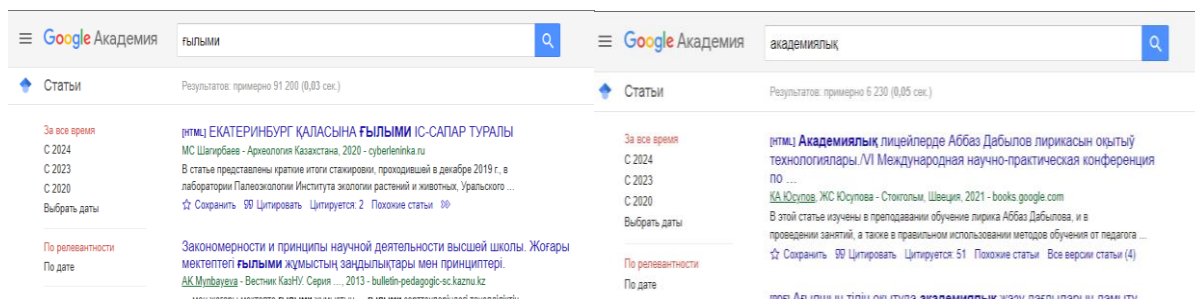
Ғылыми жоба шеңберінде негізгі ұғымдарды анықтау

Қазақ тілінің академиялық жазбаша корпусын әзірлеу жобасы аясында жүргізілген ғылыми зерттеулер барысында бірқатар терминдерді анықтап алуымыз қажет болды. Академиялық корпус әзірлеу ісі қазақ тіл білімінде алғаш қолға алынып отырғандықтан, тақырыпқа қатысты бірқатар негізгі және өзектес терминдерді нақтылап, талқылау маңызды. Ғылыми жобаның мақсаты корпусының академиялық түрін әзірлеу болғандықтан, жоба аясында жүргізілетін ғылыми зерттеулерде *академиялық* атрибут сөзі негізгі ұғым болып табылады. Бұл – қазақ тілінің академиялық жазбаша корпусын әзірлеу аясында жүргізілетін ғылыми зерттеулеріміздің басты кілт сөзі және негізгі терминдік тіркестерді жасаудағы тірек сөз.

Қазіргі қазақ тілді ғылыми ортада *академиялық* сөзінің қолданысы аз, оған қарағанда *ғылыми* сөзінің жұмсалыуы көп. *Ғылыми* және *академиялық* атрибут сөздерін салыстыра алуымыздың себебі мынада: бұл екі ұғымның мағынасы қолданыста бір-біріне жуықтасып кетеді, кейде тіпті екі сөз бірінің орнына бірі жұмсалады. *Академиялық* сөзінің қолданысы білім беру саласындағы соңғы өзгерістермен, атап айтқанда біртұтас еуропалық кеңістікке

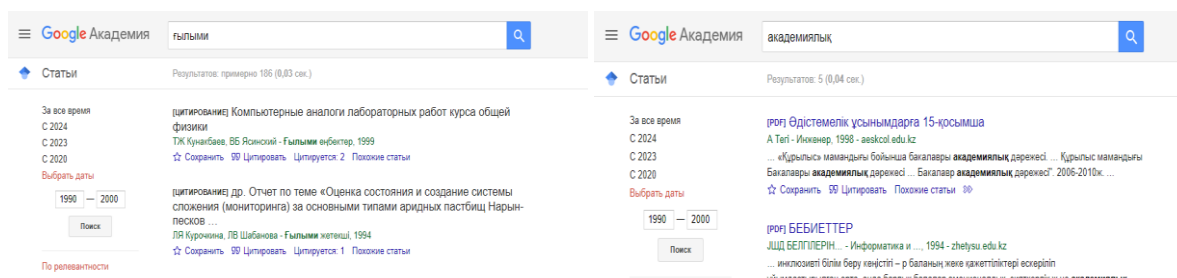
енудің Болон процесімен жоғары оқу орындарында академиялық ұтқырлық, академиялық адалдық, академиялық алмасу, сондай-ақ академиялық жазу сияқты жаңашылдықтардың енгізілуімен, өзектені түсті. Орта білім беру оқу орындарында оқушылардың *академиялық эссе* жазуы оқыту үдерісіне енгізілді. Осындай жаңа істердің іске асуымен *академиялық ұтқырлық, академиялық алмасу, академиялық орта, академиялық жазу/академиялық жазылым, академиялық дискурс* т.с.с. тіркестер қолданысқа енді. Бізді ғылыми зерттеуіміздің бағытына сәйкес *академиялық жазу/академиялық жазылым, академиялық мәтін, академиялық тіл, академиялық қазақ тілі* сөз тіркестерінің қолданыстары қызықтырады. Қазіргі қазақ тілді ғылыми айналыста *академиялық жазу/академиялық жазылым, академиялық хат, академиялық эссе* тіркестері жиі қолданылады, бірақ *академиялық тіл, академиялық қазақ тілі, академиялық мәтін, академиялық дискурс, академиялық жұмыс* ұғымдары қолданыста өте сирек ұшырасады деп ойлаймыз. Академиялық мақсаттарға арналған қазақ тілі немесе қазақ тілін академиялық мақсатта оқыту қолданыстары кездеспейді деуге болады. Олармен салыстырғанда ғылыми жазу, ғылым тілі/ғылыми тіл, ғылыми мәтін, ғылыми дискурс, ғылыми жұмыс ұғымдары өте жиі кездеседі. Сондай-ақ академиялық атрибут сөзімен келген мұндай сөздер мен сөз тіркестері қазақ тілді ғылыми ортаға соңғы онжылдықтарда енді және осы уақыттан бастап жиі қолданыла бастады деп болжамдаймыз. Ойымыз мен болжамымыздың дұрыстығын тексеру үшін Google Scholar платформасында зерттеу жасап көрдік. Google Scholar-ды таңдауымыздың себебі, ол тек академиялық еңбектерге іздеу жасайтын платформа болып табылады. Бұл біздің зерттеуіміздің мақсатына сай келеді.

Іздеуге алдымен *ғылыми* сөзін салдық. Барлық уақыт кезеңдерінде іздеу нәтижесі бұл сөзге қатысты шамамен 91 200 қолданысты тапты. Ал *академиялық* сөзінің қолданыс саны барлық уақыт кезеңімен іздегенде шамамен 6230 болды (Сурет 1). Салыстырып қарасақ, *ғылыми* сөзінің қолданысы *академиялық* сөзінің қолданылу жиілігіне қарағанда шамамен 15 есеге артық. (іздеу күні – 29.12.2024)



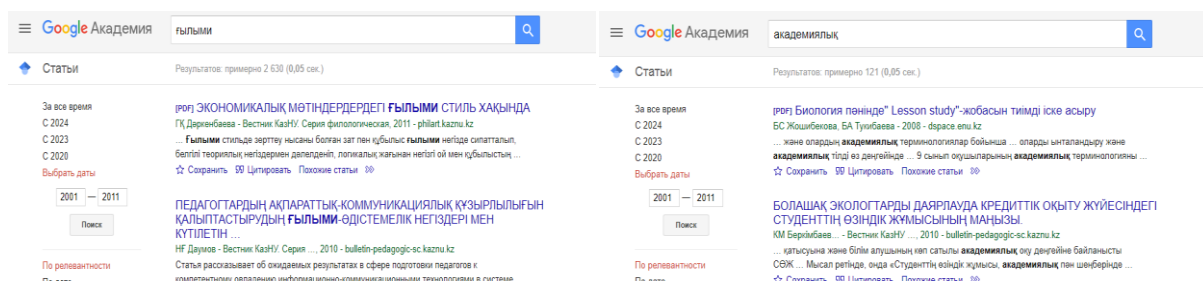
Сурет 1. Ғылыми және академиялық сөздерінің шамамен қолданылу жиілігі

Ғылыми және *академиялық* сөздері 1990 жылдан бергі онжылдықтарда қаншалықты жиі қолданылды, қолданыс жиілігінде айырма бар ма деген сұрағымызға зерттеу нәтижелері мыналарды көрсетті. *Ғылыми* сөзі 1990-2000 жылдары аралығында шамамен 186 рет қолданылса, *академиялық* сөзі дәл осы онжылдықта шамамен 5 рет ғана қолданылған (Сурет 2).

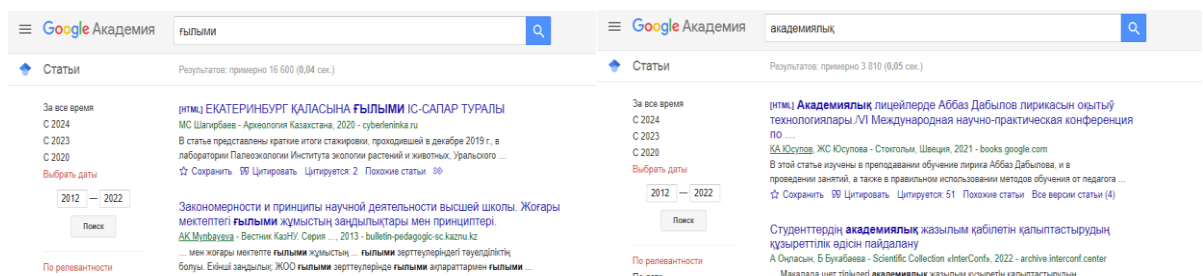


Сурет 2. Ғылыми және академиялық сөздерінің қолданылу жиілігі (1990-2000)

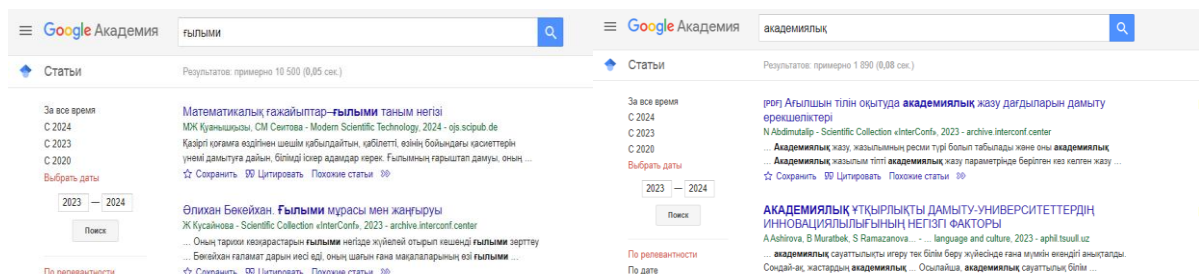
Кейінгі онжылдықтарда да *ғылыми* сөзінің қолданылу жиілігі *академиялық* сөзінің қолданылу жиілігінен бірнеше есеге артқан. 2001-2011 жылдар аралығында *ғылыми* сөзінің қолданысы 2630-ға, онан кейінгі онжылдықта (2012-2022) 16600-ге жеткен. Соңғы бір жылдың өзінде (2023-2024) *ғылыми* сөзінің қолданысы 10500 болды. Мұнымен салыстырғанда *академиялық* сөзінің дәл осы онжылдықтардағы қолданылу жиілігі әлдеқайда төмен: оның саны 2001-2011 жылдары шамамен 121 болса, 2012-2022 жылдары шамамен 31,5 есеге дейін артып 3810-ге жеткен. Соңғы бір жылда (2023-2024) *академиялық* сөзінің қолданылу жиілігі шамамен 1890 болды (Суреттер 3, 4, 5).



Сурет 3. *Ғылыми* және *академиялық* сөздерінің шамамен қолданылу жиілігі (2001-2011)



Сурет 4. *Ғылыми* және *академиялық* сөздерінің шамамен қолданылу жиілігі (2012-2022)



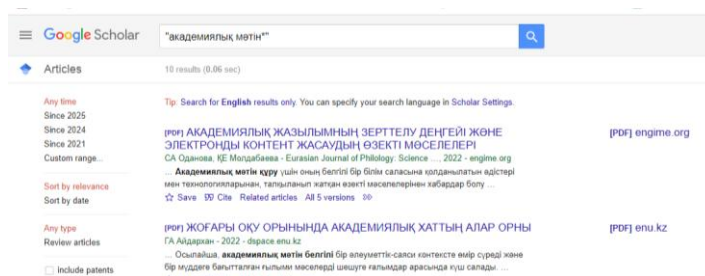
Сурет 5. *Ғылыми* және *академиялық* сөздерінің шамамен қолданылу жиілігі (2023-2024)

Бұл зерттеу екі сөзге қатысты да 1990 жылдан 2024 жылға дейін қолданылуында өсу болғанын көрсетеді және *академиялық* сөзінің XX ғасырдың 90 жылдары танымал термин болмағанын, кейінгі 2010 жылдардан бері ғана бұл сөзді жұмсау өзектеніп түскенін көрсетіп отыр. Сонда жалпы алғанда *ғылыми* сөзі 1990-2024 жылдар арасында шамамен 29916 рет қолданылған болса, *академиялық* сөзі дәл осы уақыт аралығында шамамен 5826 рет қолданылған. Бұл *академиялық* атрибутының *ғылыми* атрибут сөзіне қарағанда шамамен бес есеге кем жұмсалғанын көрсетеді.

Сөздердің қолданыс жиілігіне зерттеу біздің алғашқы болжамымыздың дұрыстығын көрсетті. Яғни *ғылыми* сөзінің қазақ ғылымында қолданылу жиілігі өте жоғары болса, *академиялық* сөзінің қолданылу жиілігі төмен және ол кейінгі 2010 жылдардан бастап қана өзектеніп бастаған.

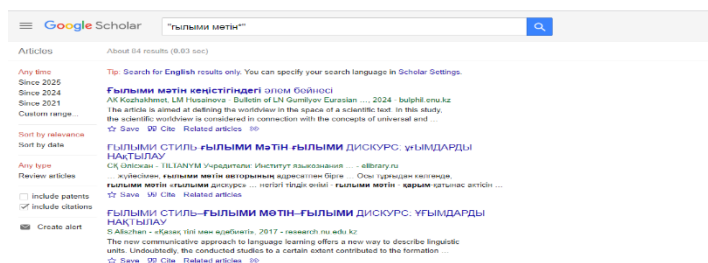
Академиялық сөзінің тірек сыңар болуымен қазақ тілінде бірқатар терминдік тіркестер жасалады. Солардың ішінде біздің зерттеуімізге қатысты негізгі ұғымдардың, атап айтқанда *академиялық мәтін*, *академиялық қазақ тілі*, *қазақ тілін академиялық мақсатта оқыту*,

академиялық жазу/академиялық жазылым, академиялық корпус сөз тіркестерінің, қолданылуына назар аударайық. *Академиялық мәтін* ұғымының қолданысы іздеуде қатысы жоқ қолданыстарды алып шыққандықтан, іздеу аясын нақтыладық. Нәтижесінде барлық уақыт кезеңінде *академиялық мәтін* тіркесі шамамен 10 рет ғана қолданылған (Сурет 6).



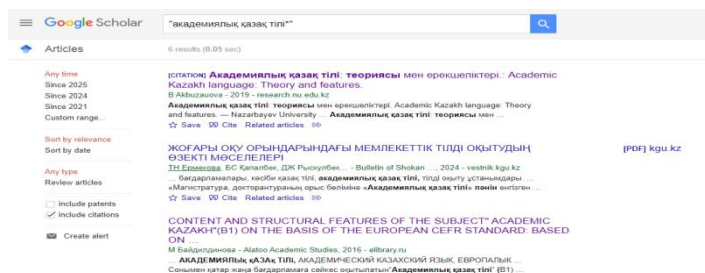
Сурет 6. Академиялық мәтін сөз тіркесінің шамамен қолданылу жиілігі

Бұл сөз тіркесімен салыстырғанда *ғылыми мәтін* сөз тіркесі шамамен сегіз есе артық қолданылған (Сурет 7).



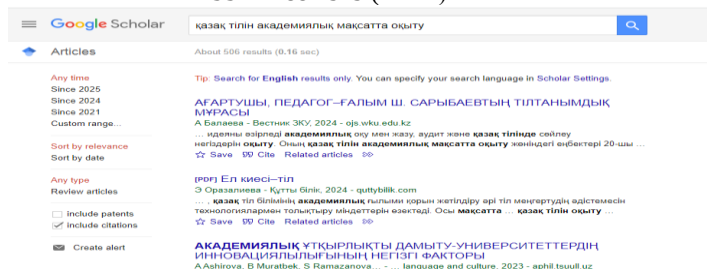
Сурет 7. Ғылыми мәтін сөз тіркесінің шамамен қолданылу жиілігі

Академиялық қазақ тілі тіркесін іздеуде де ұғымға қатысы жоқ қолданыстарды есепке алмау мақсатында нақтыладық. Бұл тіркес шамамен 6 рет ғана қолданылған, соның ішінде біреуі бір мақаланың қайталануы болса, қалған бесеуі академиялық қазақ тілін оқытуға қатысты қолданылған (Сурет 8).



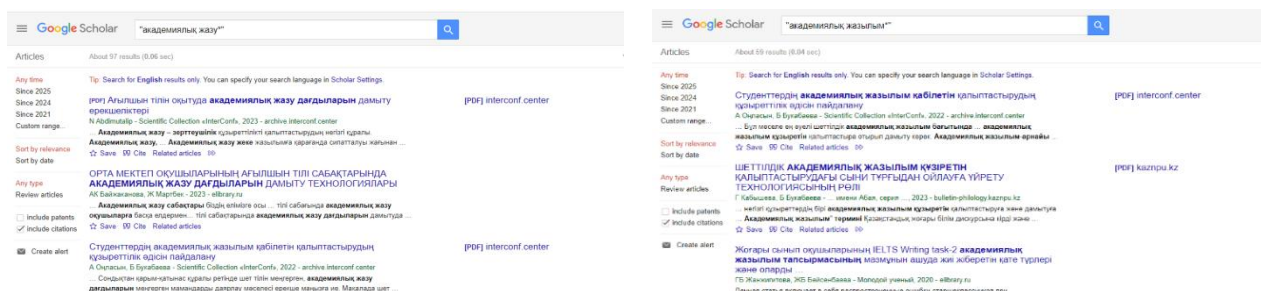
Сурет 8. Академиялық қазақ тілі сөз тіркесінің шамамен қолданылу жиілігі

Қазақ тілін академиялық мақсатта оқыту тіркесі тіпті қолданыста жоқтың қасы десек болады. Уақытқа шектеу қоймай жалпы іздегеннің өзінде тіркес бір-ақ рет кездесті (Сурет 9). (іздеу күні – 03.01.2024)



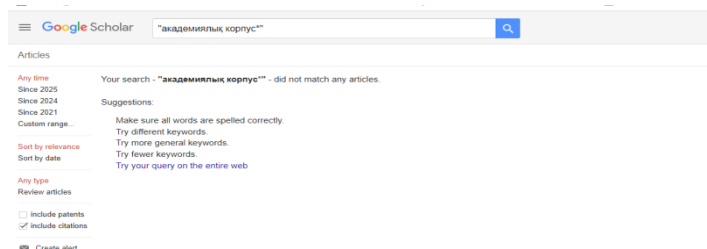
Сурет 9. Қазақ тілін академиялық мақсатта оқыту тіркесінің шамамен қолданылу жиілігі

Академиялық жазу мен академиялық жазылым сөз тіркестерін баламалы терминдер ретінде іздегенде, іздеу нәтижесі академиялық жазу шамамен 97 рет, ал академиялық жазылым шамамен 59 рет қолданылғанын көрсетті (Сурет 10). (іздеу күні - 05.01.2025)



Сурет 10. Академиялық жазу/академиялық жазылым сөз тіркестерінің шамамен қолданылу жиілігі

Академиялық корпус терминіне іздеу еш нәтиже бермеді (Сурет 11). Бұл біздің зерттеу тақырыбымыздың негізгі термині қазақ тілді ғылымда қолданыста жоқ екенін көрсетеді.



Сурет 11. Академиялық корпус сөз тіркесінің шамамен қолданылу жиілігі

Сонымен, Google Scholar платформасында жүргізілген зерттеу академиялық жазбаша қазақ тілінің корпусын әзірлеу жобасы тақырыбындағы ғылыми зерттеулердің негізгі терминдері мен терминдік тіркестерінің кейбірінің қолданыста аз, кейбірінің өте аз кездесетінін және енді бірінің әлі де ғылыми айналымға енгенін көрсетті. Бұл корпус лингвистикасы саласында қазақ тілінде ғылыми зерттеулер жүргізудің және корпустар құру ісінің өзектілігін және маңыздылығын көрсетеді. Бұл істер аталған саладағы басты термин сөздер мен сөз тіркестерінің қолданысын арттырады және қазақ тілді ғылыми қолданыста жоқ ұғымдар мен олардың аталуы терминдердің ғылыми айналысқа енуін мүмкін етеді.

Академиялық корпустар құрудың шетелдік тәжірибесі

Көпшілікке ұсынылған академиялық корпустардың ішінде саны жағынан ағылшын тілінен құрылған корпустар көп. Олардың ішінде ертеректе 1990 жылдары құрылғандары да, соңғы онжылдықтарда құрылғандары да бар. Репрезентативті корпустар да, репрезентативті емес корпустар да кездеседі. Сөз саны жағынан көлемі үлкен және шағын түрлері де бар. Мысалы, өткен ғасырдың 90 жылдарының аяғында құрылған Ағылшын тіл иелерінің

эсселерінің Лувен корпусы (Louvain Corpus of Native English Essays, LOCNESS) [3] көлемі жағынан шағын корпус болып есептеледі. Бұл корпус - ағылшын тілінде жазылған эсселерді жинақтаған корпус. Лувен корпусы Британдық оқушылардың 60 209 сөзден тұратын А деңгейіндегі эсселерінен, Британдық университет студенттерінің 95 695 сөзден тұратын эсселерінен және Американдық университет студенттерінің 168 400 сөзден тұратын эсселерінен тұрады. Сонда бұл корпустың сөздердің жалпы саны – 324 304 сөз [4].

Ағылшын тілін үйренушінің халықаралық корпусы (International Corpus of Learner English, ICLE) ортадан жоғары және жоғары деңгейлердегі тіл үйренушілердің жазбаша эсселерінің корпусы болып табылады. Бұл корпустың негізін Лувен университетінің зерттеушісі Сильвиан Грейджер қалаған және бұл корпус халықаралық деңгейдегі көптеген университеттердің 30 жылға жуық бірлескен қызметінің нәтижесінде жасалған жұмыс болып табылады [5]. 2002 жылы бірінші басылымы және 2009 жылы екінші басылымы шыққаннан бері корпус халықаралық ғылыми жобалардың ауқымында кеңінен қолданылып келеді. Корпус түрлі ғылыми зерттеулерді жүргізуге негіз болған. Осы корпусқа сүйеніп жүргізілген зерттеулердің нәтижесінде көптеген магистрлік және кандидаттық диссертациялар орындалған. Корпустың бүгінгі күнге дейін үш нұсқасы жарық көрді. Әр нұсқасында корпустар тілдік фон және сөз саны жағынан жетілдіріліп, дамытылған. 2002 жылғы бірінші нұсқасы 11 ана тілдік фон үйренушілер қолданған 2,5 миллион сөзден тұрса, 2009 жылғы екінші нұсқасы тілдік фон (саны - 16) жағынан да, сөздер саны жағынан да (сөз саны - 3,7 миллион) көлемдірек болды. Үшінші нұсқа 25 ана тілден алынған деректерді қамтиды және сөздің саны 5,5 миллионнан тұрады. Соңғы нұсқасының тағы бір артықшылығы, алдыңғы нұсқалардың екеуі де CD-ROM дискілерінде таратылған болса, соңғы нұсқасы жаңа вебке негізделген ICLEv3 интерфейсте орналастырылған. Бұл аталған корпустардың атауында *академиялық* атрибуты болмағанын және тек оқушылардың эсселері мен студенттердің жазбаша жұмыстарын қамтығанын байқауға болады.

Академиялық корпустардың бірқатарын ағылшын тілін шет тілі ретінде меңгертудің TOEFL, TESOL бағдарламалары шеңберінде жазылған мәтіндерден тұратын корпустар құрайды. Солардың бірі – TOEFL 2000 Ауызша және жазбаша академиялық тіл корпусы (TOEFL 2000 Spoken and Written Academic Language Corpus, (T2K-SWAL) [6]. Бұл корпусты құрушылар TESOL мамандарына қажет университетте қолданылатын барлық ауызша және жазбаша регистрлердің толық лингвистикалық сипаттамасын білу мақсатын көздегендіктерін айтады [6]. Осы мақсатқа сәйкес бұл корпус АҚШ университеттерінде қолданылатын ауызша және жазбаша регистрлердің толық спектрін, атап айтқанда сыныптағы оқытуды, кеңсе сағаттарын, оқу топтарын, оқулықтарды, қамтыған. Сондай-ақ негізгі академиялық пәндерді, атап айтқанда гуманитарлық ғылымдар мен жаратылыстану ғылымдарының пәндерін, сондай-ақ академиялық деңгейлерден төменгі бөлім, жоғары бөлім және бітірушілер түрлерін қамтыған.

Академиялық және ғылыми ортада кеңінен танымал болған академиялық корпустардың бірі – Британдық академиялық жазбаша ағылшын тілінің корпусы (BAWE – British Academic Written English Corpus) [2]. Британдық академиялық жазбаша ағылшын тілі (BAWE) – Ұлыбританиядағы университеттерде жазылған студенттердің жазбаша жұмыстарының басын біріктіріп ұсынып отырған академиялық корпус. Бұл корпус өнер және гуманитарлық ғылымдар, әлеуметтік ғылымдар, өмір туралы ғылымдар және физика ғылымдарынан бакалавриат пен магистратура деңгейлерінде жазылған британдық академиялық ағылшын тілінің үлгілерін жинақтаған. Бүкіл корпус көлемі 500-5000 сөз аралығындағы студенттердің 2761 жазбаша жұмысынан тұрады. BAWE корпусының басқа академиялық корпустардан ерекшелігі – ондағы мәтіндердің кәсіби тұрғыдан өңделмеген және сараптаудан өтпеген жазбаша мәтіндер болуы. Мысалы, корпусқа магистрлік диссертациялар кірмеген, себебі олардың көлемі жазбаша жұмыстарға қарағанда көлемдірек болған. Сондай-ақ бастысы диссертациялар өңдеу, қайта жазу және рецензиялау процестерінен өткендіктен, жарияланған нұсқаға тым жақын деп есептеліп, енгізілмеген. Корпустың репрезентативтілігі академиялық оқудың әртүрлі кезеңдерінен төрт кезеңді: негізінен бакалавриат

бағдарламасын және соңғы төртінші жылы магистратура болып келетін кейбір төрт жылдық бакалавриат бағдарламасын, сондай-ақ әртүрлі пәндерден төрт тобын: өнер және гуманитарлық ғылымдар, жаратылыстану ғылымдары, физика ғылымдары және әлеуметтік ғылымдарды таңдап алуынан көрінеді.

Мичиган корпусы (Michigan Corpus of Upper-level Student Papers) да студенттердің жазбаша жұмыстарынан құрылған корпус. Корпусты Мичиган университетінің Ағылшын тілі институтының зерттеушілері әзірлеген. Корпусты әзірлеушілер Мичиган университетінің тек А бағалауын алған бакалавриаттың соңғы және магистратураның алғашқы курстарында жазылған студенттердің жазбаша жұмыстарын іріктеп алған. Корпусты әзірлеушілердің айтуынша, корпус академиялық жазудағы пән айырмашылықтарын көруге, белгілі бір сөз не сөз тіркесі қолданысының әртүрлі тәсілдерін бақылауға және жазбаша жұмыстардың түрлерінің, мәселен дәйексөздермен немесе қатаң аргументпен жазылған жұмыстардың, ерекшеліктерін табуға мүмкіндік береді [7]. Бұл корпусты түрлі саланың пәндерін және жазбаша жұмыстардың бірнеше түрлерін қамтыған, академиялық мәтін түрлерін кеңінен ұсынған репрезентативті корпус және құрылымдық тұрғыдан жақсы жасалған корпус деп бағалауға болады.

Академиялық ағылшын тілінің Оксфорд корпусы (Oxford Corpus of Academic English (OCAE) [8] көлемі жағынан ең үлкен корпус болып табылады. Қазіргі көлемі 85 миллион сөзден тұрады. Осы корпустың негізінде Тіл үйренушіге арналған академиялық ағылшын тілінің Оксфорд сөздігі (Oxford Learner's Dictionary of Academic English) жасалған. Сонымен қатар аталған корпус студенттердің академиялық жазу дағдыларын дамытуға қажетті басқа да құралдарды ұсынады. Корпус жүздеген академиялық қолданыстың жазбаларын, нақты академиялық мәтіндерге негізделген толық сөйлем мысалдарын, референстер беттерін, сондай-ақ Oxford Academic iWriter интерактивті жазу тьюториалын қамтиды.

Кембридж академиялық ағылшын тілінің корпусы (CAEC - Cambridge Academic English Corpus) [9] да студенттердің жазбаша және ауызша жұмыстарының мәтіндерінен құрастырылған корпус. Корпусқа АҚШ пен Ұлыбританияның бірқатар институттарынан бакалавриат және жоғары оқу орнынан кейінгі деңгейдегі жазбаша және ауызша академиялық тілден жиналған мәтіндер кірген. Академиялық мәтіндер дәрістерден, семинарлардан, студенттердің презентацияларынан, журналдардан, эсселерден және оқулықтардан іріктелген.

Қарап отырсақ, пайдаланудағы академиялық корпустардың басым бөлігі ағылшын тілінен және шетел зерттеушілерінің тобы әзірлеген. Сонымен қатар, зерттеуіміздің алдыңғы бөлімінде атап өткеніміздей, ағылшын тіліндегі академиялық жазбаша корпустар негізінен студенттердің, магистранттардың жазбаша жұмыстарының және ағылшын тілін үйренушілердің эсселерінің мәтіндерінен құрылған. Ол корпустардың көлемі де әртүрлі: бірнеше мыңдаған сөзден бастап миллиондаған сөзден асып түседі.

Академиялық қазақ тілінің корпусын әзірлеудің өзектілігі және маңыздылығы

Академиялық жазбаша қазақ тілінің корпусын әзірлеу алғаш рет қолға алынғалы отыр. Жоғарыда жүргізген зерттеуіміздің нәтижесінің *академиялық корпус* терминдік тіркесінің әлі ғылыми қолданыста жоқ екендігін және басқа да негізгі ұғымдардың ғылыми айналымда аз кездесетінін көрсетуі бұл тақырыпта ғылыми зерттеулер жүргізудің өзектілігін және қазақ тілінде академиялық корпус құрудың қажеттілігін дәлелдейді. Корпустар құру, оның ішінде академиялық корпустар құру ісі, қазақ тіл білімінде корпус лингвистикасы, компьютер лингвистикасы сынды қолданбалы тіл ғылымының салаларын дамытуда және гуманитарлық ғылымдарды цифрландыру ісін арттыруда өзекті. Академиялық қазақ тілінің жазбаша корпусы қазақ тілінің корпус лингвистикасы саласының дамуына қосылатын үлес болумен қатар, қазақ тілінің ғылыми және академиялық әлеуеті мен функционалдық мүмкіндіктерін көрсететін электронды оқу-зерттеу және ақпараттық-танымдық құрал болмақ.

Қазақ тіл ғылымы корпустар құру ісі жағынан Орталық Азия кеңістігінде көш бастап келе жатқанымен, халықаралық деңгейде көп артта қалып келеді. Бұл салада қазіргі уақытта қалың көпшілік пайдалана алатын ашық қолжетімді екі корпусты, атап айтқанда Алматы

қазақ тілі корпусын және Қазақ тілінің ұлттық корпусын ғана айта аламыз. Бұлардан басқа қалың көпшілікке пайдалануға ашық қолжетімді емес Л.Н. Гумилев атындағы Еуразия ұлттық университетінің, Назарбаев университетінің ғалымдары мен зерттеушілері әзірлеген қазақ тілінің жазбаша корпусы және сөйлеу тілінің корпусы [10] бар. Біздің ғылыми зерттеу жобамыздың аясында әзірленетін академиялық жазбаша қазақ тілінің корпусы қазақ тілінде жарияланған, жалпыға қолжетімді академиялық жазбаша мәтіндер жиынтығы арқылы қазақ тіліндегі қазір қолда бар корпустар ресурстарын толықтыру мақсатын көздейді, сол арқылы қазақ тілді корпустардың санын және түрлілігін арттыруға үлес қосады.

Академиялық корпус құрудың өзектілігі мен қажеттілігі тағы мынадан көрінеді. Академиялық корпустардың білім беру мен оқыту ісінде кеңінен қолданылатынын халықаралық тәжірибе көрсетіп отыр. Осы тұрғыдан қазақ тілінің академиялық корпусы елімізде және шет елдерде қазақ тілін үйрену мен үйретуге қажетті электронды әдістемелік құрал бола алады. Корпус қазақ тілін мемлекеттік тіл ретінде оқытуда және басқа ұлт өкілдеріне шет тілі ретінде меңгертуде мемлекеттік тілдің лингводидактикалық және лингвоқолданбалы базасын жасау ісіне үлес болмақ. Академиялық корпус сонымен қатар білім берудің жоғарыдан кейінгі, жоғары және кәсіптік оқу орындарында академиялық қазақ тілі пәнін оқыту ісінде қажет. Бұл мақсатта ол қазақ тілін академиялық мақсатта оқытуда пайдаланылатын цифрлық платформа болады.

Қазақ тілінің академиялық жазбаша корпусын әзірлеу мынадай мүмкіндіктерге жол ашуымен де маңызды:

Біріншіден, корпус құру ісінің аясында лексикографиялық зерттеулер жүргізу, сөздіктер құру қоса жүреді. Ағылшын тіліндегі белгілі көптеген сөздіктер (Коллинз, Вебстер, Макмиллан және т.б.) үлкен лингвистикалық корпустардың негізінде жасалған. Біздің ғылыми жобамыздың аясында да сөздік құру ісі қолға алынады. Бұл лексикографтарға академиялық жазудағы сөздердің, сөз тіркестерінің, фразеологизмдердің үлкен мәтіндер көлеміндегі қолданыстарын тексеруге, өлшеуге және талдауға мүмкіндік береді;

Екіншіден, ғылыми жоба нәтижесінде жасалатын корпус арқылы академиялық қазақ тілінің лексикалық қабаты жайлы, академиялық салалар мен пәндерде қолданылатын сөздер мен фразалардың салыстырмалы жиіліктері туралы дәлді деректер алуға мүмкіндік болады;

Үшіншіден, академиялық қазақ тілінің корпусы әртүрлі академиялық салалардағы пәндерде академиялық мәтіндердің жанрлары мен регистрлерінің кең ауқымында кездесетін белгілі бір сөздер мен сөздер тобын, контекстегі кілт сөздерді, конкорданстарды, коллокацияларды анықтауға және зерттеуге мүмкіндік береді;

Төртіншіден, корпус лингвистикасы тілдердің табиғи қолданыстағы грамматикасын, атап айтқанда белгілі бір грамматикалық құбылыстардың бір-бірімен үйлесімділігін, зерттейтіндіктен, біз әзірлейтін корпус арқылы академиялық мәтіндердегі грамматикалық тұлғалардың тіркесімділіктерін зерттеу мүмкін болады;

Бесіншіден, корпус арқылы тұтас мәтіндерді зерттеуге жол ашылады. Мысалы, жобадағы мәтін корпусының көмегімен академиялық мәтіннің функционалды стилін оның статистикалық сипаттамалары, мысалы, сөз бен сөйлемнің орташа ұзындығы, сөздердің қалыпты тіркестерінің саны және т.б., арқылы анықтауға болады;

Алтыншыдан, ғылыми жоба нәтижесінде ұсынылатын корпусты, жоғарыда атап өткеніміздей, лингводидактикада, яғни тілдерді оқытуда, оның ішінде қазақ тілін оқытуда қолдануға болады. Тілді үйретушіге нені оқыту керектігін білу үшін оқытылатын тіл туралы нақты сандық мәліметтер қажет, мысалы, жиі кездесетін лексиканың құрамы, академиялық сөздер тізімі, белгілі бір грамматикалық түзілімдерді қолдану ықтималдығы және т.б. Дәл осы сандық мәліметтерді академиялық корпус бере алады.

ҚОРЫТЫНДЫ

Зерттеу жұмысының мақсаты ұлттық тілде академиялық жазбаша корпус әзірлеп, ұсынудың қазіргі уақытта өзекті әрі маңызды екенін көрсету болды. Қазіргі қолда бар қазақ

тілінде құрылған корпустар қатарына академиялық корпусты қосу корпустардың санын арттырумен қатар, түрлілік жағынан да байытады. Академиялық жазбаша қазақ тілінің корпусы қазақ тілінің жазба тілдегі академиялық әлеуеті мен қолданыс мүмкіндіктерін айғақтайтын электронды ресурс болуымен де құнды болмақ. Мұндай корпустардың негізінде сан түрлі ғылыми зерттеулер орындалады. Мақаламызда атап өткеніміздей, корпустар, соның ішінде академиялық корпустар, сан түрлі ғылыми зерттеулердің, докторлық диссертациялардың орындалуына негіз болуда. Ендеше, қазақ тілінің академиялық корпусының негізінде де лингвистикада ғана емес, әдебиет, аударма ісі, тілдерді оқыту және тағы басқа салаларда корпустық негізделген ғылыми зерттеулер жүргізу мүмкін болады.

Қазақ тілінде академиялық жазбаша корпус әзірлеу аясында ғылыми зерттеулер жүргізу барысында бірқатар ұғымдардың мәні мен қолданысы өзектенеді. Зерттеу барысында тақырыпқа қатысты негізгі терминдердің қазіргі ғылыми айналымда қаншалықты жиі немесе сирек қолданылатыны айқындалды. Кейбір терминдердің соңғы онжылдықтарда белсендірек қолданыла бастағаны байқалды. Академиялық корпуспен байланысты қолданылатын негізгі терминдердің қазақ тілді ғылыми қолданыста сирек жұмсалуды немесе мүлдем кездеспеуі бұл тақырыптың өзектілігін, корпус құрудың және осы тақырыпта ғылыми зерттеулер жүргізудің қажеттілігін көрсетеді.

Академиялық корпустар лингвистикалық зерттеулер жасауда, тілдерді оқытуда, жалпы лингводидактикада кеңінен қолданылады. Академиялық корпустар құрылған тілдерде осы корпустарды пайдаланып көптеген сапалық және сандық ғылыми зерттеулер жасалды және жасалуда. Ендеше, қазақ тілінде академиялық мәтіндердің корпусын жасау өзекті. Ол цифрлық технологияны гуманитария ғылымына енгізуде, сол арқылы гуманитарлық ғылым салаларындағы зерттеулердегі тілдік деректер мен құралдарды, гуманитарлық зерттеулер мен компьютерлік зерттеулерді ұштастыра отырып пәнаралық зерттеулер жүргізуді дамытуда маңызды.

ҚАРЖЫЛАНДЫРУ ТУРАЛЫ АҚПАРАТ

Бұл зерттеуді Қазақстан Республикасы Ғылым және жоғары білім министрлігінің Ғылым комитеті қаржыландырды (Грант №AP23488585 «Цифрлық гуманитарлық ғылымдар: Академиялық қазақ тілінің корпусын әзірлеу»).

ӘДЕБИЕТТЕР ТІЗІМІ

- 1 Sinclair J. Trust the text: Language, corpus and discourse. – Routledge, 2004. – P. 51-66.
- 2 Alsop S., Nesi H. Issues in the development of the British Academic Written English (BAWE) corpus // Corpora. – 2009. – Т. 4. – №. 1. – P. 71-83. <http://dx.doi.org/10.3366/E1749503209000227> (қаралған мерзімі: 08.11.2024).
- 3 Granger S., Tyson S. Connector usage in the English essay writing of native and non-native EFL speakers of English // World Englishes. – 1996. – Т. 15. – №. 1. – P. 19–29.
- 4 Granger S. The computer learner corpus: a versatile new source of data for SLA research // Learner English on computer. – Routledge, 2014. – P. 3-18.
- 5 Granger S. The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research // Tesol Quarterly. – 2003. – Т. 37. – №. 3. – P. 538-546. <https://doi.org/10.2307/3588404> (қаралған мерзімі: 02.11.2024)
- 6 Biber D., Conrad S., Reppen R., Byrd P., and Helt M. Speaking and writing in the university: A multidimensional comparison // TESOL quarterly. – 2002. – Т. 36. – №. 1. – P. 9-48.
- 7 Michigan Corpus of Upper-level Student Papers <https://sites.lsa.umich.edu/eresources/2020/05/29/micusp-corpus-of-written-academic-papers/> (қаралған мерзімі: 17.11.2024).
- 8 Oxford Corpus of Academic English (OCAE). <https://www.sketchengine.eu/oxford-corpus-of-academic-english/> (қаралған мерзімі: 17.11.2024).

- 9 CAEC: Cambridge Academic English Corpus. <https://www.sketchengine.eu/cambridge-academic-english-corpus/> (қаралған мерзімі: 14.11.2024).
- 10 Mussakhojayeva S., Khassanov Y., Varol H. A. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus // INTERSPEECH. – 2022. – P. 1367-1371.

Материал 25.04.2025 баспаға түсті

Академический корпус и актуальность и значимость его разработки на казахском языке

Г.А. Сарсеке¹, Б.Е. Акмагамбетова², Ж.К. Сентирбек¹

¹Евразийский национальный университет им. Л.Н. Гумилева, Астана, 010000, Республика Казахстан

²Северо-Казахстанский университет им. М. Козыбаева, Петропавловск, 150000, Республика Казахстан

В статье представлены результаты научных исследований по академическим корпусам и разработке академического письменного корпуса казахского языка. Целью научного исследования является обоснование актуальности и важности разработки письменного корпуса академического казахского языка. В статье рассматриваются состояние современной казахской корпусной лингвистики и вклад нашего научного исследования в это направление, зарубежный опыт создания академических корпусов, а также актуальность и важность создания корпуса академического казахского.

В данной работе выявлены основные термины по теме и раскрыты их использования в современных научных исследованиях в результате исследований, проведенных в рамках разработки корпуса академического письменного казахского языка. Проведен анализ зарубежного опыта создания академических корпусов, в том числе академических корпусов, созданных на английском языке, по типу, содержанию, структуре, объему и репрезентативности.

Результаты исследования могут быть использованы в качестве теоретических и практических материалов в области корпусной лингвистики и компьютерной лингвистики казахского языка, а также при преподавании этих дисциплин в высших учебных заведениях. Разработка и презентация корпуса академического письменного казахского языка будет способствовать повышению научного потенциала и прикладных возможностей национального языка, широкому продвижению казахского языка в международном пространстве и ускорению цифровизации гуманитарных наук. Сам академический корпус и научные исследования, проведенные по данной теме, позволяют укрепить концепцию академического казахского языка, определить академическую лексику и всесторонне изучить академические тексты. Способствует преподаванию казахского языка для академических целей.

Ключевые слова: академический, академический корпус, академический текст, казахский язык, письменный корпус, обучение для академической цели, цифровизация, гуманитарные науки

СПИСОК ЛИТЕРАТУРЫ

- 1 Sinclair J. Trust the text: Language, corpus and discourse. – Routledge, 2004. – P. 51-66.
- 2 Alsop S., Nesi H. Issues in the development of the British Academic Written English (BAWE) corpus // Corpora. – 2009. – Т. 4. – №. 1. – P. 71-83. <http://dx.doi.org/10.3366/E1749503209000227> (дата обращения: 08.11.2024).
- 3 Granger S., Tyson S. Connector usage in the English essay writing of native and non-native EFL speakers of English // World Englishes. – 1996. – Т. 15. – №. 1. – P. 19-29.

- 4 Granger S. The computer learner corpus: a versatile new source of data for SLA research // *Learner English on computer*. – Routledge, 2014. – P. 3-18.
- 5 Granger S. The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research // *Tesol Quarterly*. – 2003. – Т. 37. – №. 3. – P. 538-546. <https://doi.org/10.2307/3588404> (дата обращения: 02.11.2024).
- 6 Biber D., Conrad S., Reppen R., Byrd P., Helt M. Speaking and writing in the university: A multidimensional comparison // *TESOL quarterly*. – 2002. – Т. 36. – №. 1. – P. 9-48.
- 7 Michigan Corpus of Upper-level Student Papers <https://sites.lsa.umich.edu/eresources/2020/05/29/micusp-corpus-of-written-academic-papers/> (дата обращения: 17.11.2024).
- 8 Oxford Corpus of Academic English (OCAE). <https://www.sketchengine.eu/oxford-corpus-of-academic-english/> (дата обращения: 17.11.2024).
- 9 CAEC: Cambridge Academic English Corpus. <https://www.sketchengine.eu/cambridge-academic-english-corpus/> (дата обращения: 14.11.2024).
- 10 Mussakhojayeva S., Khassanov Y., Varol H. A. KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus // *INTER_SPEECH*. – 2022. – P. 1367-1371.

Материал поступил в редакцию журнала 25.04.2025

Academic corpus and the relevance and significance of its development in the Kazakh language

G. Sarseke¹, B. Akmagambetova², Zh. Sentirbek¹

¹L.N. Gumilyov Eurasian National University, Astana, 010000, Republic of Kazakhstan

²M. Kozybayev North Kazakhstan University, Petropavlovsk, 150000, Republic of Kazakhstan

This article presents the findings of a study on academic corpora and the development of a written corpus of the academic Kazakh language. The primary aim is to substantiate the relevance and necessity of creating an academic written corpus in Kazakh. The paper examines the current state of corpus linguistics in Kazakhstan, outlines the contribution of this research to the field, and highlights international experiences in developing academic corpora. It emphasizes the practical and theoretical importance of establishing such a resource for the Kazakh language.

Within the framework of the project, key concepts related to academic corpora were identified and their applicability within Kazakh linguistics explored. The article provides quantitative data on the frequency of specific terms relevant to the formation of the academic corpus. It also includes a comparative analysis of academic corpora in English, focusing on their types, content, structure, size, and representativeness.

The study offers both theoretical insights for the fields of corpus and computational linguistics in Kazakhstan, and practical materials for teaching these subjects in higher education institutions. The creation and implementation of an academic written Kazakh corpus is positioned as a significant contribution to enhancing the scientific and applied potential of the national language, promoting its global presence, and advancing the digitalization of the humanities.

Key words: academic, academic corpus, academic text, Kazakh language, written corpus, teaching for academic purposes, digitization, humanities.

REFERENCES

- 1 Sinclair, J. (2004). *Trust the text: Language, corpus and discourse*. Routledge, 51-66.
- 2 Alsop, S., & Nesi, H. (2009). Issues in the development of the British Academic Written English (BAWE) corpus. *Corpora*, 4(1), 71-83. <http://dx.doi.org/10.3366/E1749503209000227> (date of access: 08.11.2024).
- 3 Granger, S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World Englishes*, 15(1), 17-27.

- 4 Granger, S. (2014). The computer learner corpus: a versatile new source of data for SLA research. In *Learner English on computer* (pp. 3-18). Routledge.
- 5 Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, 37(3), 538-546. <https://doi.org/10.2307/3588404> (date of access: 02.11.2024).
- 6 Biber, D., Conrad, S., Reppen, R., Byrd, P., and Helt, M. (2002). Speaking and writing in the university: A multidimensional comparison. *Tesol Quarterly*, 36(1), 9-48.
- 7 Michigan Corpus of Upper-level Student Papers <https://sites.lsa.umich.edu/eresources/2020/05/29/micusp-corpus-of-written-academic-papers/> (date of access: 17.11.2024).
- 8 Oxford Corpus of Academic English (OCAE). <https://www.sketchengine.eu/oxford-corpus-of-academic-english/> (date of access: 17.11.2024).
- 9 CAEC: Cambridge Academic English Corpus. <https://www.sketchengine.eu/cambridge-academic-english-corpus/> (date of access: 14.11.2024).
- 10 Mussakhojayeva, S., Khassanov, Y., and Varol, H. A. (2022). KSC2: An Industrial-Scale Open-Source Kazakh Speech Corpus. In *INTERSPEECH* (pp. 1367-1371).

Received: 25.04.2025

МРНТИ 16.01.11

DOI: [10.59102/kufil/2025/iss2pp227-242](https://doi.org/10.59102/kufil/2025/iss2pp227-242)

Д.А. Худайбердина¹, Л.С. Байманова², А.Ж. Сарсембаева³

¹Казахский национальный университет имени аль-Фараби, Алматы, 050000, Республика Казахстан

²Кокшетауский университет имени Ш.Уалиханова, Кокшетау, 020000, Республика Казахстан

³Казахская национальная академия имени Т. Жургенова, Алматы, 050000, Республика Казахстан

ОСОБЕННОСТИ ФОРМИРОВАНИЯ БИЛИНГВАЛЬНОЙ ЭЛИТАРНОЙ ЛИЧНОСТИ В ИСТОРИКО-КУЛЬТУРНОМ КОНТЕКСТЕ КАЗАХСТАНА

В статье исследуется феномен формирования билингвальной элитарной личности в историко-культурном контексте Казахстана на примере творческой биографии Мурата Ауэзова. Цель исследования – определить ключевые особенности элитарной билингвальной личности, проанализировать влияние семьи, историко-культурных и образовательных факторов на процесс её формирования. В исследовании учитывается влияние двуязычия на культурную и когнитивную компетентность, а также на развитие профессиональной и социальной идентичности. Анализируется значение билингвизма для сохранения культурных норм, роли образовательной среды в становлении элитарной личности и влияния двуязычия на когнитивные способности в период взросления. В рамках исследования применяются методы когнитивного, дискурсивного анализа и историко-культурного подхода, основанного на описании биографических данных и социально-культурных традиций. Подчеркивается важность билингвальной образовательной среды для формирования элитарной личности и условий развития когнитивных способностей. Практическая значимость результатов заключается в разработке конкретных рекомендаций по формированию образовательных стратегий и программ в области филологии и социокультуры.

Ключевые слова: билингвизм, элитарная личность, историко-культурный контекст, когнитивная компетенция, дискурсивный анализ.