

Қымбат Берікұлы Слямбеков^{1*}, Меруерт Ануаровна Имангазина^{1*}автор-корреспондент, докторант, Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Қазақстан, Алматы қ., ORCID: 0000-0002-9731-0448 E-mail: zaqpost@gmail.com²философия докторы (PhD), Ахмет Байтұрсынұлы атындағы Тіл білімі институты, Қазақстан, Алматы қ., ORCID: 0000-0001-8473-2812 E-mail: meruyertimangazina@mail.ru**ТІЛ БІЛІМІ САЛАСЫ БОЙЫНША ҚАЗАҚША АКАДЕМИЯЛЫҚ СӨЗДЕР ТІЗІМІН (AWL) КОРПУСТЫҚ НЕГІЗДЕ ӘЗІРЛЕУ**

Аңдатпа. Мақалада заманауи ғылыми стильдің қазіргі ұйымдасу прототипіне академиялық жазылым мәтіндерінде жиі қолданылатын сөзқолданыстарды анықтау, соның ішінде тілі білімі бойынша академиялық мәтіндерге тән грамматикалық құралдар мен лексика-грамматикалық құрылымдар сандық және сапалық талдау әдістері арқылы анықтауға ұмтылыс жасалған. Жеке саланы айтпағанда, жалпы академиялық мәтінге тән сөздер тізімі лексикалық минимум ретінде қазақ ғылымында әлі күнге дейін жарық көрмеген. Бұл мәселе зерттеуіміздің мақсатын былайша айқындап берді: зерттеу мақсаты – қазақ тіліндегі академиялық мәтіндерге, соның ішінде лингвистикаға қатысты зерттеу мақалаларына тән базалық сөздік қорды анықтау. Осы мақсат жолында академиялық жазылым мәтіндерінің сөзтізімін жасақтауға бағытталған отандық және шетелдік іргелі зерттеулер, тәжірибелер қарастырылды. Зерттеу материалы ретінде тіл білімі бағытындағы ҚРҒЖБССҚК ұсынған басылымдардағы 6 журналдың 30 нөмірінен алынған 236 ғылыми мақалаға негізделген жалпы көлемі 1 057 000 сөзқолданысты құрайтын «Лингвистикалық зерттеулер корпусы» жасақталды. Корпустық талдау және жиілік талдау әдістері арқылы сөзтізімге енетін ең жиі қолданылған 1000, 200 және 50 тілдік бірлік сұрыптап алынып, бұл тілдік бірліктердің қазақша академиялық жазылым мәтіндерінің қандай сипаты мен ерекшелігін, академиялық мәтіндердің қандай қағидасын қамтамасыз етуге қызмет ететіні талданды. Сұрыптап алынған тілдік бірліктер негізінде академиялық жазылым мәтіндерінің жиілік сөз тізімі жасақталды. Мұндай сөз тізімі тіл білімі бағытындағы академиялық жазылым мәтіндерінің өзіне тән өзгешеліктерін тануға, олардың қызметі мен мағыналық түрленімін айқындауға мүмкіндік береді. Сондай-ақ бұл зерттеу арқылы «қазақ тілі – ғылым тілі» екенін нақты тілдік фактілер арқылы дәлелдеуге деген ұмтылыс та бар. Зерттеудің ғылыми жаңалығы қазақ тіл білімі саласындағы ғылыми мақалалар негізінде арнайы салалық корпус құрастырылып, академиялық дискурста жиі қолданылатын тілдік бірліктер мен олардың тұрақты тіркесімдерінің сандық және сапалық тұрғыдан сипатталуымен айқындалады. Практикалық маңызы да айқын: нәтижелер академиялық жазылымды оқытуға, қазақша AWL/AVL үлгісімен сөздік жасауға және NLP міндеттеріне пайдаланылуы мүмкін.

Тірек сөздер: академиялық жазылым; ғылыми мәтін; корпус; сөздер тізімі; тілдік бірліктер; лексика-грамматикалық минимум.

Қаржыландыру көзі: Мақала Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің AP27511671 «ЖОО-да академиялық жазылым дағдыларын интеграциялық әдіс негізінде әдістемелік құралдар әзірлеу арқылы қалыптастыру» атты гранттық қаржыландыру жобасы аясында әзірленді.

Сілтеме жасау үшін: Слямбеков Қ.Б., Имангазина М.А. Тіл білімі саласы бойынша қазақша академиялық сөздер тізімін (AWL) корпустық негізде әзірлеу. *Tiltanyum*, 2026. №3 (103). 230-245-бб.

DOI: <https://doi.org/10.55491/2411-6076-2026-3-230-245>

Қымбат Берікұлы Слямбеков^{1*}, Меруерт Ануаровна Имангазина^{1*}автор-корреспондент, докторант, Институт языкознания имени Ахмета Байтурсынулы, Казахстан, г. Алматы, ORCID: 0000-0002-9731-0448 E-mail: zaqpost@gmail.com²доктор философии (PhD), Институт языкознания имени Ахмета Байтурсынулы, Казахстан, г. Алматы, ORCID: 0000-0001-8473-2812 E-mail: meruyertimangazina@mail.ru**РАЗРАБОТКА СПИСКА КАЗАХСКИХ АКАДЕМИЧЕСКИХ СЛОВ (AWL) ПО ЛИНГВИСТИКЕ НА КОРПУСНОЙ ОСНОВЕ**

Аннотация. В статье предпринята попытка выявить часто употребляемые языковые единицы, характерные для современного прототипа организации научного стиля, представленного в текстах академического письма, в том числе грамматические средства и лексико-грамматические конструкции, свойственные академическим текстам по языкознанию, с использованием методов количественного и качественного анализа. В казахстанской научной практике до настоящего времени не опубликован список слов, характерных для общеакадемического текста, не говоря уже о специализированных списках по отдельным научным направлениям. Эта проблема определила цель нашего исследования следующим образом: цель исследования – определить базовый словарный запас, характерный

для академических текстов на казахском языке, в том числе научных статей, относящихся к лингвистике. Для этой цели были рассмотрены отечественные и зарубежные фундаментальные исследования, практики, направленные на формирование словосочетаний текстов академического письма. В качестве материала исследования был создан «Корпус лингвистических исследований», включающий 236 научных статей, отобранных из 30 выпусков 6 журналов, рекомендованных Комитетом по обеспечению качества в сфере науки и высшего образования Республики Казахстан по направлению «Языкознание». Общий объем корпуса составляет 1 057 000 словоупотреблений. С помощью методов корпусного и частотного анализа были отобраны наиболее часто употребляемые 1000, 200 и 50 языковых единиц, вошедших в список слов. Далее было проанализировано, какие особенности и характеристики казахскоязычных текстов академического письма отражают данные языковые единицы и обеспечение каких принципов академического текста они способствуют. На основе отобранных языковых единиц был составлен частотный словарь текстов академического письма. Такой словарь позволяет выявить специфические особенности академических текстов по лингвистике, а также определить функции и семантические вариации употребления этих языковых единиц. Кроме того, исследование направлено на обоснование положения о том, что казахский язык является языком науки, на основе конкретных языковых фактов. Научная новизна исследования заключается в создании специализированного отраслевого корпуса на основе научных статей по казахскому языкознанию, а также в количественном и качественном анализе языковых единиц и их устойчивых сочетаний, наиболее часто употребляющихся в академическом дискурсе. Практическая значимость исследования также очевидна: полученные результаты могут быть использованы при обучении академическому письму, разработке словаря казахского языка по модели AWL/AVL, а также в задачах обработки естественного языка (NLP).

Ключевые слова: академическое письмо; научный текст; корпус; список слов; языковые единицы; лексико-грамматический минимум.

Источник финансирования: Статья подготовлена в рамках проекта грантового финансирования Комитета науки Министерства науки и высшего образования Республики Казахстан АР27511671 «Формирование навыков академического письма в вузах посредством разработки методических материалов на основе интеграционного подхода».

Для цитирования: Слямбеков К.Б., Имангазина М.А. Разработка списка казахских академических слов (AWL) по лингвистике на корпусной основе. *Tiltanyim*, 2026. №3 (103). С. 230-245. (на каз. яз.)

DOI: <https://doi.org/10.55491/2411-6076-2026-3-230-245>

Kymbat Slyambekov^{1*}, Meruyert Imangazina

^{1*}Corresponding Author, Doctoral Student, Akhmet Baitursynuly Institute of Linguistics, Kazakhstan, Almaty, ORCID: 0000-0002-9731-0448 E-mail: zaqpost@gmail.com

²Doctor of Philosophy (PhD), Akhmet Baitursynuly Institute of Linguistics, Kazakhstan, Almaty, ORCID: 0000-0001-8473-2812 E-mail: meruyertimangazina@mail.ru

CORPUS-BASED DEVELOPMENT OF A KAZAKH ACADEMIC WORD LIST (AWL) IN THE FIELD OF LINGUISTICS

Abstract. The article attempts to identify frequently used linguistic expressions characteristic of the contemporary organizational prototype of scientific style as represented in academic writing texts, including grammatical devices and lexico-grammatical constructions typical of academic texts in linguistics, using quantitative and qualitative methods of analysis. A list of words characteristic of general academic texts, let alone those specific to individual disciplines, has not yet been published in Kazakhstani scholarship. This issue determined the aim of our study as follows: the aim of the research is to identify the core vocabulary characteristic of academic texts in the Kazakh language, including research articles in the field of linguistics. To achieve this aim, fundamental domestic and international studies and practices aimed at compiling word lists for academic writing texts were reviewed. As the research material, the “Linguistic Research Corpus” was developed, comprising 236 research articles selected from 30 issues of 6 journals recommended by the Committee for Quality Assurance in Science and Higher Education of the Republic of Kazakhstan in the field of linguistics. The total size of the corpus is 1,057,000 word tokens. Using corpus and frequency analysis methods, the 1,000, 200, and 50 most frequently occurring linguistic units included in the word list were identified. These linguistic units were subsequently analyzed to determine what characteristics and features of Kazakh academic writing texts they reflect and which principles of academic writing they contribute to fulfilling. Based on the selected linguistic units, a frequency word list of academic writing texts was compiled. Such a word list makes it possible to identify the specific features of academic texts in the field of linguistics and to determine the functions and semantic variations of these linguistic units in use. Furthermore, the study seeks to demonstrate, through concrete linguistic evidence, that “the Kazakh language is a language of science.” The scientific novelty of the study lies in the development of a specialized disciplinary corpus based on research articles in Kazakh linguistics, as well as in the quantitative and qualitative description of linguistic units frequently used in academic discourse and their recurrent combinations. The practical significance of the study is also evident: the findings can be used in teaching academic writing, in developing a Kazakh academic word list based on the AWL/AVL model, and in natural language processing (NLP) tasks.

Keywords: academic writing; scientific text; corpus; word list; linguistic units; lexical-grammatical minimum.

Source of financing: The article was prepared within the framework of the grant financing project of the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan IRN AP27511671 “Developing academic writing skills in universities through the creation of methodological materials based on an integrative approach”.

For citation: Slyambekov, K., Imangazina, M. Corpus-Based Development of a Kazakh Academic Word List (AWL) in the Field of Linguistics. *Tiltanyim*, 2026. No.3 (103). Pp. 230-245. (in Kazakh)

DOI: <https://doi.org/10.55491/2411-6076-2026-3-230-245>

Кіріспе

Академиялық жазылымды меңгеру барлық сала студенттері мен зерттеуші мамандар үшін маңызды дағды саналады. Себебі академиялық мәтін – зерттеуші қозғап отырған мәселенің өзектілігін, зерттеу барысының қалай жүзеге асқанын және қол жеткізген нәтижелерін ғылыми қауымға жеткізудің негізгі тетіктерінің бірі. Мұнда ғылыми мәтіннің құрылымы ғана емес, мәтіннің өн бойында терминдер мен академиялық жазылымға тән басқа да тұрақты қолданыстарды, тіл бірліктерін ұтымды қолданудың да маңызы зор. Себебі ғылыми мәтінді құраушы тілдік бірліктер мәтін авторының білім-білігін ғана емес, ғылыми стильді қаншалықты меңгергенін өзі зерттеп отырған тақырыбына қаншалықты дендеп енгенін айқындай түсетін лексикалық қабат саналады. Екіншіден, белгілі бір кезеңдегі ғылыми мәтіндерде жиі ұшырасқан қолданыстар сол кезеңде аталған ғылым саласы үшін өзекті болған тақырыптарды, зерттеу проблемалары мен оларды шешу жолдарын да болжауға мүмкіндік береді. Сол себепті ұсынылып отырған мақалада лингвистика бағытында жазылған және соңғы екі жылда Қазақстан Республикасы Ғылым және Жоғары білім министрлігінің Ғылым және жоғары білім саласындағы сапаны қамтамасыз ету комитеті базасына енетін журналдарда жарық көрген, яғни сапалы әрі ғылыми әлеуеті жоғары деп саналатын мақала мәтіндерінің корпусы құрылып, солардың негізінде тіл біліміне қатысты зерттеу мақалаларындағы ең жиі қолданылған тіл бірліктері анықталды.

Сөздерді іріктеп алу мақсатында тек лингвистика бағытындағы зерттеу мақалаларын дереккөз ретінде таңдап алуымыздың екі негізгі себебі бар. Біріншіден, зерттеуімізде бір саланың мәселесін шешіп алуды көздеп отырмыз, кейін осы үлгіге негізінде басқа салалардың да академиялық жиілік сөздігін жасауға болады. Екіншіден, шетелдік зерттеушілер академиялық сөздер тізімін әр сала негізінде жасау әлдеқайда нақты әрі ұтымды деп есептейді. «Бірнеше корпустық зерттеу нәтижелері көрсеткендей, түрлі ғылым салаларына ортақ және практикалық тұрғыдан тиімді болатындай ауқымды академиялық лексика қорын бөліп көрсету мүмкін емес» (Durrant, 2016).

Ағылшын тіліндегі академиялық сөздер тізіміне жиі сын айтылуының бір себебі олардың бірнеше сала деректерін негізге алуында болып отырғанын да естен шығармаған жөн. Сондай-ақ дереккөздерді таңдауда, яғни корпус жасақтауда назар аударуымыз керек тағы бір мәселе – ондағы академиялық тілдің барынша жаңа, яғни қазіргі қазақ тілі деректері болуы, сондай-ақ жиі оқылатын, қаралатын және қолданылатын (сілтеме берілетін) мәтіндердің қамтылуы. Бұл тұрғыдан ҚР Ғылым және жоғары білім министрлігінің Ғылым және жоғары білім саласындағы сапаны қамтамасыз ету комитетінің жоғары рейтингілі журналдарының мақала қаралымы анық көрініп тұратындықтан және бұл басылымдарда бұрын соңды жарық көрмеген тың зерттеулер жарияланып отыратындықтан, жоғарыда қойылып отырған талапқа толық сай келеді дей аламыз.

Зерттеу мақалалары автордың индивидуалды стилі мен дүниетанымының, ғылыми танымының жемісі болғандықтан мәтіннің бұл түрінде нақты бір қалыпқа салуға келе бермейтін оралымдар мен сөз қолданыстардың болуы заңды. Алайда біз нысанға алып отырған мәтіндердің барлығы бір стиль мен бір жанрдың туындысы әрі журнал талаптары негізінде құрылымдалған мәтіндер болғандықтан академиялық стильге жат сөздер мен сөз тіркестері, мысалдар мен иллюстративті материалдарда болмаса, аса жиі көрініс таппайды. Мұндағы автор шығармашылығын көркем әдебиет, публицистикалық немесе ауызекі сөйлеу стилімен салыстыруға келмейді. Сондай-ақ зерттеу мақалалары Committee on Publication Ethics ұсынған этика талаптарына сәйкес жазылады. Біз жинаған корпус материалдарын осы тұрғыдан да нақты дереккөз ретінде пайдалануға әбден лайық деп санаймыз.

Материал және әдістер

Зерттеу жұмысы бірнеше кезеңді қамтыды. Ең алдымен, академиялық сөздер тізімін жасау тәжірибесіне қатысты отандық және шетелдік зерттеулер талданып, теориялық негізі анықталды. Екінші кезекте, жиілік сөздікті түзуге қажет дереккөздер іріктеп алынып, тіл білімі бағытындағы ғылыми мақалалардан тұратын шағын академиялық корпус түзілді. Осы сәттен бастап жұмыстың негізгі кезеңіне кірістік. Корпус деректері Google Colab бағдарламасында талданып, ең жиі кездесетін 1000 қолданыс сұрыпталды және лексика-грамматикалық тұрғыдан компоненттік талдау жасалды. Алынған 1000 сөзқолданыс ішінен мақала авторлары академиялық мәтіндерде тұрақты түрде кездесетін және қолданушыға ғылыми мәтін түзуде негіз болатын 200 сөзқолданыс бөлек жинақталып, академиялық мәтінге тән сөз-тізім құрастырылды.

Зерттеу мақсаты тіл білімі саласына қатысты академиялық сөздікқорды сұрыптау болғандықтан, зерттеу дереккөзі ретінде еліміздегі лингвистика және филология бағытындағы журналдар тандап алынды. Соның ішінде филология серияларындағы тек тіл біліміне қатысты материалдар ғана корпусқа енді. Нәтижесінде зерттеу материалдары ретінде «Тілтаным» журналының 2025-2026 жылғы мақалалары (5 нөмірі, 57 мақала), сонымен қатар аталған кезеңдегі «ҚазҰУ Хабаршысы. Филология сериясы» (5 нөмірі, 43 мақала), «ЕҰУ Хабаршысы. Филология сериясы» (5 нөмірі, 32 мақала), Қарағанды Университетінің Хабаршысы. Филология сериясы» (5 нөмірі, 24 мақала), Абылай хан атындағы ҚазХҚ және ӨТУ Хабаршысы. Филология сериясы» (5 нөмірі, 48 мақала) «Ш.Уәлиханов атындағы Көкшетау университетінің хабаршысы. Филология сериясында» (5 нөмірі, 32 мақала) жарияланған тіл біліміне қатысты мақалалары іріктеп алынды. Барлығы 6 журналдың 30 нөмірінен 1 млн 57 мың сөзқолданыстан тұратын қазақ тіліндегі тіл білімі саласына қатысты 236 мақаладан тұратын корпус жасақталды. Айта кететін тағы бір мәселе, мақалаларды корпусқа қоспас бұрын мақала авторы жайлы мәліметтер, орысша және ағылшынша аңдатпалар мен кілт сөздер, сондай-ақ әдебиеттер мен references бөлімі түгелдей алып тасталды. Себебі бұл ақпараттар қазақ тіліндегі академиялық мәтіндердің сөздер жиілігін шығаруда қажет емес, басы артық ақпараттық жүк арқалайды.

Корпус мәтіндерінің статистикасын шығару мен сандық талдау әдістерін жүзеге асыруға Qazcorpus (qazcorpus.kz) жобасының «Сөзжиілік модулі» қолданылды. Бұл құрал – қазақ тілінің жиілік сөздіктерін жасауға қолданылып жүрген, морфологиялық анализаторы қосылған, сөзді бөлек бір лемма деп танып, оның мәтін ішіндегі жиілігін автоматты шығарып беретін ашық қолданыстағы лингвистикалық құрал. Құралдың сөз жиілігін анықтаудағы тиімділігі 80% деп көрсетілген (<https://qazcorpus.kz/>). Негізгі проблемасы – омоним сөздерді ажырата алмайтынында. Дегенмен біздің зерттеуімізге бұл тұсы қиындық тудырмайды, өйткені ғылыми тіл мен академиялық жазылым мәтіндерінде тілдік бірліктер әдетте бір мағынада ғана жұмсалады да, омонимияға жол бермейді.

Академиялық мәтіндердің сөз жиілігін анықтауға Сөзжиілік модулімен бірге Python бағдарламалау тілімен жұмыс істеуге арналған Google Colab онлайн-сервисі пайдаланылды. Мәтіндер корпусы Google Colab ортасына жүктеліп, Python бағдарламалау тілінің көмегімен өңделді. Мәтінді оқу, тазалау, сөздерге бөлу және сөз жиілігін шығару үшін *pandas*, *re*, *Counter*, *python-docx*, *openpyxl* кітапханалары қолданылды. Бұл сервис әсіресе белгілі бір тірек сөздер айналасындағы тұрақты қолданыстарды жинауға, жеке қарастыруға мүмкіндік берді. Ал Python кодтарын жазуға ChatGPT үлкен тіл моделі қолданылды.

Автоматты өңдеу нәтижесінде алынған тілдік бірліктер қолмен тексеріліп жөнделді. Өйткені компьютерлік талдау тек тіркестерді механикалық түрде жинап береді де, оның синтаксистік және семантикалық тіркесімін ажырата алмайды. Сондықтан алғашқы сандық талдау әдісінің материалдарына сапалық талдау әдісі қолданылды.

Мақалалардағы жиі кездесетін сөздерді іріктеу барысында олардың қатарына академиялық стильге жатпайтын бірліктер еніп кетуі мүмкін деген қауіп болғаны белгілі. Алайда әлемдік тәжірибеге сай академиялық сөздер тізімінің басым бөлігін шын мәнінде академиялық жазылымға тән бірліктер құрайды. Бұл тұрасында 3,5 миллион сөзқолданыстан тұратын корпус негізінде академиялық мәтін сөздер тізімін жасап шыққан Коксхед былай дейді: «Ағылшын тілінің академиялық сөздер тізімінде 570 сөз тобы бар, олар академиялық мәтіндердегі сөздердің

(таңбалауыштардың) жалпы санының шамамен 10,0%-ын құрайды, алайда дәл сондай көлемдегі көркем әдебиет мәтіндеріндегі сөздердің жалпы санының тек 1,4%-ын құрайды. Екі түрлі мәтіндегі бұл айырмашылық тізімдегі сөздердің басым бөлігін шын мәнінде академиялық сөздер құрайтынының дәлелі болып табылады. Бұл тізім академиялық мәтіндерде жиі кездесетін сөздерді бөліп көрсету арқылы академиялық мақсаттары бар студенттерге қай сөздерді меңгеру қажет екенін анықтап бере алады» (Coxhead, 2000: 213).

Академиялық мәтіндер корпусын түзу мәселесі қазақ ғылымында азды-көпті сөз болып жүргенімен, нақты шешімін тапты дей алмаймыз. Қазақ ғылымында академиялық корпус терминін алғаш қолданған Г.Ә. Сәрсек, Б.Е. Ақмағамбетова, Ж.Қ. Сентірбек мақаласында бұл ұғымға мынандай сипаттама береді: «Академиялық корпустар академиялық салалардың әртүрлі пәндеріндегі тілдің қолданысы, нақтылай айтқанда академиялық жазбаша мәтіндерде тілдің қалай қолданылатыны туралы түсінік береді. Академиялық корпустар академиялық мәтіндердің түрлі жанрлары мен регистрлерінен жиналған үлкен көлемі арқылы академиялық жазбаша және сөйлеу тілінің табиғи қалпын тануға көмектеседі. Корпус деректері арқылы академиялық мәтіндердің басқа мәтін түрлерінен ерекшеленетін өзіндік белгілері мен сипаттамаларын алуға болады» (Сәрсек, Ақмағамбетова, Сентірбек, 2025: 213).

Отандық зерттеушілер: «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» (Сәрсек, Ақмағамбетова, Сентірбек, 2025: 214), – деп, бұл мәселенің өзектілігін дәл көрсеткен. Алайда мақала авторларының «Компьютерлік технологиялардың мүмкіндіктерін пайдалана отырып, қазіргі қазақ тілінің академиялық жазбаша корпусын құру – аса маңызды әрі қажетті іс. Отандық тәжірибеде академиялық корпус(тар) жоқ» (Сәрсек, Ақмағамбетова, Сентірбек, 2025: 214) деген пікірін дәл қазіргі уақытта сол қалпында қайталай алмаймыз. Себебі қазіргі таңда Қазақ тілінің ұлттық корпусында «Ғылыми мәтіндер ішкорпусы» жасақталған. Ішкорпус «академиялық мәтіндер ішкорпусы» деп аталмағанымен, контенті бұл талапқа толық сай келеді. Корпус құрастырушылары бұл тұста қазақ тілінде «ғылыми» сөзінің «академиялық» сөзіне қарағанда жиірек қолданылатынын да ескерген болуы керек. Мәселен, соңғы деректерге сай «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» (Сәрсек, Ақмағамбетова, Сентірбек, 2025: 218).

 Г.О. Шалшықбаева, К.К. Дүйсекова, Р.М. Бактиярова Авиация кәсібі дискурсының конститутивтік ерекшеліктері	
Мәтін авторы	Г.О. Шалшықбаева, К.К. Дүйсекова, Р.М. Бактиярова
Автор жынысы	әйел
Автордың дәрежесі	1) Абылай хан атындағы Қазақ халықаралық қатынастар және әлем тілдері университетінің докторанты, 2) Л.Н. Гумилев атындағы Еуразия ұлттық университеті ф.ғ.д., профессор
Мәтін атауы	Авиация кәсібі дискурсының конститутивтік ерекшеліктері
Жарық көрген уақыты	2025 ж.
Журнал атауы	Еуразия гуманитарлық университеті Хабаршысы
Журнал саны	№1
Беті	54-65 бб.
Мәтін дереккөзі	https://ojs.egi.kz/BULLETIN/article/view/818
Doi сілтемесі	https://doi.org/10.55808/1999-4214.2025-1.06
Аудитория жасы	ересек
Ішкорпус түрі	Ғылыми мәтіндер корпусы
Мәтін стилі	ғылыми
Мәтін типі	мақала
Сөзқолданыс саны	2482
Мәтін формасы	жазба
Мәтін графикасы	кирил
Ғылым саласы	тіл білімі
Белгілеуші	А. Досанова

Сурет 1 – Қазақ тілінің ұлттық корпусы «Ғылыми мәтіндер ішкорпусының» метабелгіленімі

Рисунок 1 – Метаразметка «Подкорпуса научных текстов»

Национального корпуса казахского языка

Figure 1 – Metadata Annotation of the “Subcorpus of Scientific Texts” of the National Corpus of the Kazakh Language

Ғылыми мәтіндер ішкорпусының іздеу жүйесінде *мәтін авторы, ғылым саласы, автор жынысы, мәтін атауы, мәтін типі, жарық көрген уақыты* параметрлері бойынша іздеуге болады. Ішкорпустың метабелгіленімдері мақала жайлы толық ақпарат алуға мүмкіндік береді. Метабелгіленімде мақаладағы сөзқолданыс саны да көрсетіледі (Сурет 1).

Ғылыми мәтіндер ішкорпусымен қатар «Цифрлық гуманитарлық ғылымдар: Академиялық қазақ тілінің корпусын әзірлеу» ғылыми жобасы аясында отандық зерттеушілер тобы философия, тарих, дінтану, археология және этнология, шығыстану, теология, түркітану, шығыстану, мұражай ісі, шетел филологиясы, аударма ісі, қазақ филологиясы сынды гуманитарлық ғылымдардың деректері қамтылған ауқымды академиялық корпус жасақтайды деген болжам бар. Жоба авторларының айтуынша, Қазақ тілінің академиялық корпусында *мәтін деңгейі* (кімдерге арналғаны, яғни бакалавриат, магистратура, докторантура, профессор-оқытушылар құрамы); *пән саласы; жанры; пәні* (мамандығы); *шифры; академиялық мәтін тақырыбы; баспасы; авторлардың аты-жөні; авторлар коды* (метадеректе тіркелген); *жарияланым мерзімі; деңгейі; мәтіннің сөз саны; мәтіннің бет саны; авторлардың саны; авторлардың жынысы* (ер, әйел, аралас тип) және *корпусқа тіркелген күні* сынды он алты түрлі мәліметтен тұратын метадерек болады (Қоңыратбаева, 2025: 88). Қазақ тіліндегі академиялық мәтіндердің мұндай ауқымды базасын жасау қазақ тілін ғылым тілі ретінде өзектендіре түсетіні әрі қазақ тілінің ғылыми әлеуетін арттыру бағытындағы ізденістерге жол ашатыны айқын.

Әдебиетке шолу

Қазақ тілінің ғылым тілі ретінде қалыптасуы, соның ішінде қазақ тіл білімі бағытының академиялық жазылым мәселелерінің даму барысы жайлы сөз қозғағанда қазақ лингвистикасы жайлы ғылыми еңбектердің басында А. Байтұрсынұлы, Қ. Жұбановтардың тұрғанын ерекше атап өту қажет. А. Байтұрсынұлы қазақ тілі жайлы алғашқы ғылыми еңбектің – оқу құралының құрастырушысы болса, Қ. Жұбанов 1935 жылы Мемлекеттік терминологиялық комиссияны құрған. Тілші-ғалымдар Р. Сыздық пен Б. Әбілқасымов қазақ тіліне қатысты алғашқы академиялық стильге ие еңбек ретінде Ильминскийдің 1914 жылы жарық көрген «Самоучитель киргизского языка русским и русского языка киргизам» еңбегін атайды (Odanova, Tuleup, Moldabayeva, 2024). Тіл тарихына үңілген зерттеушілер пікіріне сүйене келе, бір ғасырдан астам тарихы бар деп шамалауға болатын қазақ лингвистикасының академиялық стилі қазіргі таңда құрамы мен құрылымына қатысты тың ізденістер жасауды қажет етеді. Қазақстандық зерттеушілер Е. Спанов, Б. Динаева мен С. Сапина өз еңбектерінде жоғары оқу орындарында академиялық жазылым мен сауаттылықты қалыптастырудың теориялық негіздерін жүйелеп шыққанымен, бұл зерттеулерде лингвистика саласындағы нақтыланған академиялық сөз тізімдері (AWL) көрініс таппаған. Қ. Молдабаева, М. Өнер, С. Оданованың зерттеу мақаласында академиялық жазылым дағдыларын студенттерге үйрету үшін заманауи әдіс-тәсілдерге жүгіну керек екенін айта келе, соның бір үлгісі ретінде академиялық жазылым дағдыларын қалыптастыруға бағытталған сандық тренажер құрудың маңызына ерекше тоқталады. Аталған авторлар тобы зерттеудің негізгі мақсатын «жиналған ақпаратқа сүйене отырып, қазақ тіліндегі академиялық жазудың электрондық мазмұнын жобалау, атап айтқанда, тілдік модельдер негізінде онлайн тілдік тренажерларды дайындау, барлық академиялық лексиканың электронды сөздігін дайындау. үш тілде, академиялық жазу дағдыларын қалыптастыратын онлайн арналарды құру» деп көрсетеді (Öner, Moldabayeva, Odanova, 2024: 153). Мұндағы академиялық сөздікқорды (academic vocabulary) бір жерге топтастыратын электронды сөздік жасау идеясы біздің мақаламыздағы академиялық сөздер тізіміне (academic word list) ұқсас екенін атап өту қажет.

Алайда аталған зерттеудің нәтижесінде авторлар ғылыми басылымдар негізінде қазақ тіліндегі академиялық жазылымның негізгі ұғымдарын сұрыптау және сипаттаумен шектелгенін байқадық, яғни екі зерттеудің зерттеу нысаны мен зерттеу тәсілдері екі түрлі екеніне көз жеткіздік. Біздің зерттеуімізде біз лингвистика бағытындағы академиялық мәтінді жазып отырған білімгер немесе мәтін авторы қолдануы құпталатын тілдік бірліктердің тізімін ұсынып, оларды қолдану жолдарын негіздеп береміз, яғни бұл зерттеуіміз академиялық мәтінді әдіснамалық тұрғыдан сипаттау бағытында емес, лингвистикалық жағынан талдау бағытында

жазылып отыр.

Ағылшын тілінде академиялық сөздер тізімін жасау «Академиялық мақсатқа арналған ағылшын тілі» (English for Academic Purposes) пәні аясында өткен ғасырдың елуінші жылдарынан бастап қолға алына бастаған. Ең алғаш West 1953 жылы ағылшын тілінде ең жиі қоланылатын 2000 сөздің тізімін жасайды және бұл тізімдегі сөздердің 76%-ы академиялық мәтіндерде кездескен (Coxhead, 2000: 213). Сондай-ақ XX ғасырдың екінші жартысында Кэмптон (Campion) мен Элли (Elley), Пранискас (Praninskas), Линн (Lynn), Гадесса (Ghadessy) сынды зерттеушілер де академиялық мәтінге тән сөздер тізімін қолмен жинап, сөздігін жасауға талпынған. Ал осы төрт зерттеуді негізге ала отырып және оларды толықтыра отырып Сюэ Гоу (Xue Guoyi) мен И. С. П. Нейшн (I. S. P. Nation) *University Word List* деген атаумен зерттеушілерге арналған сөздер тізімін әзірлеген.

Ағылшын тіліндегі ең соңғы әрі толыққанды тізім ретінде Аверил Коксхедтың 2000 жылы құрастырған академиялық сөздер тізімін (Academic Word List) атай аламыз. Бұл тізімді жасау үшін автор, ең алдымен, Академиялық корпус (Academic Corpora) түзген. Автордың бұл корпусы қолданушының үдесінен шығу үшін төрт ішкорпусқа бөлінген. Коксхед түзген 570 сөзқолданыстан тұратын академиялық сөздер тізімін сөздер тобына (word families), басқаша айтқанда, сөзжасамдық ұяға жіктеп берген. Мәселен, *maximize* сөзінің ұяластары ретінде *maximised*, *maximises*, *maximising* сынды формалары *maximisation* деген британ ағылшын нұсқасы мен американ ағылшын нұсқасындағы *maximization* сөзі де енген. Бұл тізімде нақты бір сөзжасамдық ұяның құрамына кірмейтін *nonetheless*, *so-called*, *behalf* сынды байланыстырушы шылаулар жеке орналасқан (Coxhead, 1998).

Алайда мұндағы сөздер тек жалаң тізім түрінде берілгендіктен, сөздің қолданылу жиілігі жайлы экстралингвистикалық ақпарат, реестрдегі сөздің лексикалық мағынасы, контекстік мағынасы, сөз табы, грамматикалық формасы сынды лингвистикалық ақпарат та мүлдем көрсетілмеген. Сол себепті қазіргі таңда ағылшын тіліндегі 2000 жылы жарық көрген академиялық сөздер тізіміне қатысты сыни көзқарастар жиі айтылып жүр. Бельгиядағы Ағылшын корпус лингвистикасына арналған орталық маманы Магали Пако бұл мәселеде тек жиілікке негізделген сөздер тізімін берумен шектелу жеткіліксіз екенін айтады. Ағылшын тіліндегі соңғы зерттеулерді назарға ала отырып М. Пако «Академиялық мақсатқа арналған ағылшын тілінің» (English for academic purposes) өзіндік фразеологиялық жүйесі қалыптасқанын атап өтеді және «жаңа сөздерді қалай қолдану керек екені жайлы, әсіресе оларды тізімге коллокациялық және коллигациялық ортасы туралы ақпаратпен бірге енгізу өте маңызды» деп есептейді (Paquot, 2010). М. Пако ұстанымының тағы бір ерекшелігі, зерттеуші академиялық сөздер тізімін әзірлеуде тіл үйренушілердің деректері жиналған оқу корпусындағы (Learners' corpus) мәтіндерді талдап, сол жердегі қате, орынсыз немесе стильге сай келмейтін қолданыстарды анықтау қажет деп түйеді.

Ағылшын тілінде Гарднер және Дэвис 2014 жылғы зерттеуінде СОСА корпусының 120 миллион сөзқолданыстан тұратын академиялық ішкорпусы негізінде жинаған қолданыстарын *Academic Vocabulary List* (AVL) деп атаған (Gardner, Davies, 2014). Зерттеушілер 9 саланы қамтитын ғылыми журналдар мен ғылыми-көпшілік журналдар негізінде 3014 лемманы іріктеп алған (Durrant, 2016). Бұл бағыт ағылшын тілін екінші тіл ретінде немесе шет тілі ретінде үйреніп жүрген білімгерлердің сөздік қорын арттыруды мақсат етеді.

Алайда біздің зерттеу мәселеміз, ең алдымен, қазақ тілі ана тілі болып саналатын білімгерлер мен зерттеушілерге көмекші құрал дайындау, одан кейінгі кезекте оны тіл үйретуге бейімдеп қолдану болғандықтан, корпустық зерттеу нысаны ретінде қазақ тілінде жазылған академиялық мәтіннің бір түрі – ғылыми мақалалар алынып отыр.

Нәтижелер және талқылау

Зерттеуімізде 1 млн 57 мың сөзқолданыстан құратын мәтіндер корпусынан барлығы 94 мың сөздің жиілігін шығардық. Олардың ішінде ең жиі қолданылған сөздің жиілігі – 12101 рет болса, ең азы – 1 реттен қолданылғандары. Сурет 2-де мәтіндер корпусынан ең жиі кездескен 51 тілдік бірліктің тізімі берілген. Бір айта кетерлігі, «Сөзжиілік модулі» тілдік бірліктердің морфологиялық түрленімдерін бөлек лемма деп қарастырады (сондықтан сөз саны көп).

Мәселен, «тіл» сөзінің жиілігін бөлек (8-лемма), одан туған «тілдік» сөзінің жиілігін бөлек (6-лемма) көрсетеді (Сурет 2). Сол секілді *де* етістігінің *деп* және *деген* түрленімдерін де бөлек-бөлек шығарған (14-15-леммалар).

1	Base	Word	Count	26	зертте/ет	зерттеу	1958
2	және/шл	және	12101	27	талда/ет	талдау	1951
3	мен/шл	мен	10541	28	немесе/шл	немесе	1819
4	қазақ/оним	қазақ	7863	29	қат/эт	қатар	1796
5	бұл/ес	бұл	6432	30	ғылыми/сн	ғылыми	1792
6	тіл/эт	тілдік	4549	31	тіл/эт	тілінің	1780
7	бір/ес	бір	3774	32	ол/ес	олардың	1662
8	тіл/эт	тіл	3940	33	ғана/шл	ғана	1645
9	арқылы/шл	арқылы	3739	34	пен/шл	пен	1638
10	рет/шл	ретінде	3698	35	қатысты/ет	қатысты	1560
11	да/шл	да	3479	36	тіл/эт	тіліндегі	1554
12	де/шл	де	3080	37	тіл/эт	тілінде	1504
13	сөз/эт	сөз	2972	38	сөз/эт	сөздер	1465
14	де/ет	деп	2934	39	негізгі/сн	негізгі	1457
15	де/ет	деген	2558	40	әлеуметтік/сн	әлеуметтік	1441
16	үшін/шл	үшін	2494	41	белгілі/сн	белгілі	1402
17	ұлттық/сн	ұлттық	2460	42	тек/шл	тек	1393
18	мәдени/сн	мәдени	2363	43	маңызды/мод	маңызды	1390
19	емес/ет	емес	2375	44	сияқты/мод	сияқты	1380
20	бол/ет	болады	2322	45	осы/ес	осы	1357
21	бол/ет	болып	2199	46	көрсет/ет	көрсетеді	1322
22	бар/мод	бар	2145	47	туралы/шл	туралы	1305
23	байланысты/шл	байланысты	2162	48	тіл/эт	тілі	1272
24	ол/ес	оның	2146	49	қазіргі/сн	қазіргі	1268
25	мәдени/сн	мәдени	1972	50	екі/са	екі	1263
				51	жаңа/сн	жаңа	1256

Сурет 2 – Лингвистикалық академиялық корпуста ең жиі қолданылған 50 тілдік бірліктің тізімі

Рисунок 2 – Список 50 наиболее частотных языковых единиц
в Лингвистическом академическом корпусе

Figure 2 – List of the 50 Most Frequently Used Linguistic Units in the Linguistic Academic Corpus

Сурет 2-ден байқағанымыздай, академиялық жазылым мәтіндерінде жиі қолданылатын тілдік бірліктердің ішінде жалғаулық шылаулар, септеулік шылаулар, сілтеу есімдіктері, *де*, *е* және *бол* етістіктері, *талдау*, *зерттеу*, *көрсету* сынды маркерлік етістіктер, модаль сөздер (*бар*, *мүмкін*, *маңызды*), лингвистика ғылымына тән *тіл*, *сөз* сынды зат есімдер, *тілдік*, *ұлттық*, *мәдени*, *ғылыми*, *әлеуметтік*, *негізгі* сынды қатыстық сын есімдердің қолданысы жоғары болып шыққан. Бұл бір жағынан, академиялық жазылым мәтіндерінің қазақ тілінің өзге де стиль және жанр мәтіндерімен лексика-грамматикалық құрылымы мен құралдары тұрғысынан өзара ұқсас екенін, қазақ әдеби тілінің заңды тармағы екенін көрсетеді. Өйткені басқа жиілік сөздіктер де аталмыш тілдік бірліктердің қолданысы басым екенін аңғаруға болады (Негізгі корпус мәтіндерінің жиілік сөздігі, 2025: 46). Екінші жағынан, жоғарыда көрсетілген тілдік бірліктер ғылыми ақпаратты жүйелеу, толықтыру және біріктіру қызметіне жақын құралдар болғандықтан, академиялық жазылым мәтіндері *салыстыру*, *біріктіру*, *ұқсату*, *жүйелеу* сынды адамзаттың интеллектуалдық-логикалық категорияларына басымдық беретінін, сондықтан мұндай мәтіндердің когезиялық байланысы мен мәтінтүзілімін құрайтын тілдік бірліктердің қолданысы көп болатынын көрсетеді. Сондықтан қазақша академиялық мәтіндердің сөз тізімінде аталмыш лексика-грамматикалық бірліктердің тізімі алдыңғы қатарда болуы керек деп есептейміз. Мұндағы *лексика-грамматикалық бірлік* деп отырғанымыз – белгілі бір сөз табына, яғни лексика-грамматикалық топқа жататын сөздер. Зерттеуде сөздердің леммасымен қатар грамматикалық түрленімдері арқылы көрінетін мағыналық ерекшеліктері де назарға алынған.

Біз жинаған мәтіндерде ең жиі қолданылған тілдік бірлік – *және* шылауы көбіне өзара ұқсас ұғымдарды байланыстыру және тізбектеу үшін қолданылған: *қазақ және ағылшын тілдері*, *орыс және ағылшын тілі*, *қазақ және ноғай тілдері*, *қазақ және түрік тілдері*, *теориялық және практикалық маңызы* / *аспектісі* / *жағы* / *тұрғыдан* / *негізі* / *алғышарты*, *қазақ және орыс*

тілдері, ағылшын және қытай тілі, түрік және қазақ тілдері, араб және парсы сөздері, отандық және шетелдік тіл білімі / ғалымдар / еңбектер / зерттеушілер / әдебиеттер, лексикалық және грамматикалық мағына / бірлік / минимум / форма т.б. (Сурет 3).

“және” сөзімен келген үш сөзді тіркес Қолданылу саны					
0	қазақ және ағылшын	115	21	морфологиялық және синтаксистік	21
1	нәтижелер және талқылау	77	22	айту және қаржыландыру	21
2	орыс және ағылшын	69	23	когнитивтік және прагматикалық	20
3	материал және әдістер	60	24	материалдық және рухани	20
4	қазақ және ноғай	49	25	ағылшын және қазақ	20
5	қазақ және түрік	48	26	мәдени және әлеуметтік	19
6	теориялық және практикалық	46	27	қазақ және қытай	19
7	қазақ және орыс	46	28	қазақ және саха	18
8	ағылшын және қытай	39	29	семантикалық және прагматикалық	17
9	тілдік және мәдени	37	30	тарихи және әлеуметтік	16
10	түрік және қазақ	35	31	сыртқы және ішкі	16
11	отандық және шетелдік	34	32	ішкі және сыртқы	16
12	араб және парсы	33	33	сандық және сапалық	16
13	ғылым және жоғары	32	34	когнитивтік және нейролингвистикалық	15
14	мәдени және тарихи	31	35	рухани және мәдени	14
15	тарихи және мәдени	30	36	когнитивтік және мәдени	14
16	лексикалық және грамматикалық	27	37	оң және теріс	14
17	шетелдік және отандық	27	38	құрылымдық және мағыналық	14
18	әлеуметтік және мәдени	24	39	рухани және материалдық	14
19	мәдени және тілдік	22	40	мәдени және саяси	13
20	ауызша және жазбаша	21			

Сурет 3 – Лингвистикалық академиялық корпустағы «және» сөзімен тіркесіп келген бірліктер

Рисунок 3 – Единицы, сочетающиеся со словом «және»

в Лингвистическом академическом корпусе

Figure 3 – Collocations with the Word “zhane” (and) in the Linguistic Academic Corpus

Одан бөлек, *және* шылауының академиялық мәтінде нәтиже мен процесті байланыстыру, аргументтерді үдету және толықтыру, кейде бекіту, қандай да бір маңызды аргументті актуалдау қызметтерін атқаратыны байқалды.

Нәтиже мен процесті байланыстыру: *Дегенмен кей мысалдардағы (алма ағашы) тіркестердің табиғатындағы сын есімдік мәннің болуы және оларға септік жалғауларын жалғауға мүлде болмайтындығы, ал кейбір мысалдардағы (ағаш жапырағы) меншіктік мағынаның берілуі және қосымшаның қатысуымен де қатысуынсыз қолданыла алуы, сонымен қатар, қосымшамен берілген прагматиканың қосымшасыз прагматикадан айырмашылықта келуі ғылымда толық дәлелденуі қажет құбылыс деп білеміз.*

Аргументтерді үдету және толықтыру, бекіту: *Және бұл тек түркі халықтарына ғана тән құбылыс; Және зерттеу тақырыбының өзектілігін қазіргі қоғам өміріндегі БАҚ-тың маңызды рөл атқаруымен, сондай-ақ цифрлық технологиялардың қарқынды дамуымен түсіндіреді;*

Қандай да құбылысқа маңызды аргументті актуалдау: *Және бір маңызды жайт – жалпақ бетті, қошқар мұрынды, қалың ерінді деген үлгілердің «екеуі бірігіп басқа бір сөзді анықтап, бағыныңқы қызметте жұмсалып тұрған» жоқ, сөз тіркесінің өзінен кейінгі сөздің күрделі анықтауышы сияқты жұмсалуы – қазақ тілі синтаксисінің нормасы; Және бір ден қоярлығы – осындағы күрделі сөздердің екінші сыңарына -ды жұрнағын жалғайтын болсақ (қынама белді камзол, қаршыға көзді жігіт, қаба сақалды жігіт, т.б.), керісінше, олар сөз тіркесіне айналып, бұл жұрнақтың жоғарыдағы «күрделі сөз жасауға тікелей әсері бар» деген аксиома теріске шығады.*

Жалғаулық шылаулардан кейін кездесетін бірлік – бұл есімдігі. Оның 6500-ге жуық қолданысы анықталды. Сілтеу есімдіктері де жалғаулық шылаулар секілді мәтіннің тұтастығын, когезиясын және актуализациясын қамтамасыз етеді. Бұл есімдігімен келетін мынандай паттерн бірліктер табылды (Сурет 4).

"бұл" сөзімен келген екі сөзді тіркес Қолданылу саны					
0	бұл жерде	142	20	бұл олардың	37
1	бұл зерттеу	106	21	бұл зерттеудің	36
2	бұл сөз	93	22	бұл зерттеуде	35
3	бұл тәсіл	88	23	бұл ұғым	34
4	бұл құбылыс	81	24	бұл сөздердің	33
5	бұл мақалада	67	25	бұл құбылысты	31
6	бұл әдіс	66	26	бұл кезеңде	30
7	бұл ретте	57	27	бұл әсіресе	30
8	бұл тілдік	52	28	бұл тіл	29
9	бұл бағыттағы	51	29	бұл жағдайда	28
10	бұл тіркес	50	30	бұл оның	28
11	бұл сөздер	47	31	бұл тек	28
12	бұл туралы	46	32	бұл еңбек	27
13	бұл сөздің	44	33	бұл жағдай	27
14	бұл екі	43	34	бұл тілдің	26
15	бұл өз	41	35	бұл да	25
16	бұл үдеріс	41	36	бұл мәселені	23
17	бұл қазақ	41	37	бұл бағытта	23
18	бұл мәселе	40	38	бұл бір	23
19	бұл тұрғыда	40	39	бұл үрдіс	22
20	бұл олардың	37	40	бұл мақал	20
			41	бұл зерттеулер	19

Сурет 4 – Лингвистикалық академиялық корпустағы «бұл» сөзімен тіркесіп келген құрылымдар
 Рисунок 4 – Конструкции, сочетающиеся со словом «бұл»
 в Лингвистическом академическом корпусе
 Figure 4 – Constructions with the Word “бұл” (this) in the Linguistic Academic Corpus

Бұл тілдік бірліктердің қазақша академиялық жазылым мәтіндерінің тіліне тән тұрақты қолданыстар екенін байқауға болады. Әсіресе, *бұл зерттеу*, *бұл тәсіл*, *бұл құбылыс*, *бұл әдіс*, *бұл бағытта*, *бұл тіркес*, *бұл туралы*, *бұл үдеріс*, *бұл мәселе*, *бұл тұрғыдан*, *бұл сөз*, *бұл ұғым*, *бұл жағдайда*, *бұл мәселе*, *бұл бағытта*, *бұл үрдіс* т.б. тіркестер қазақша академиялық жазылымның мәтінтұтастырушы бірліктері, референция құралдары, стильдік қолданысы деуге болады. Сондай-ақ *бұл зерттеу жұмысы*, *бұл зерттеудің өзектілігі*, *бұл зерттеудің нәтижелері*, *бұл зерттеудің мақсаты*, *бұл өз кезегінде*, *бұл бағыттағы зерттеулер*, *бұл тұрғыдан алғанда* т.б. сынды үш компонентті тіркестер академиялық жазылым мәтіндерінің метадискурсын құрайтын, яғни автордың мәтін құрылымын дұрыс ұйымдастыруына, алдыңғы және кейінгі ойын бір-бірімен байланыстыруына мүмкіндік беретін, сол арқылы оқырманға зерттеуін жүйелі таныстыруына жол ашатын стандартталған стильдік қолданыстар екенін аңғаруға болады.

Академиялық жазылым мәтіндерінде септеулік шылаулардың да қызметі ерекше екені тілдік фактілерден айқын көрінді (*арқылы*, *үшін*, *ретінде*, *байланысты*, *сияқты*, *қатысты*, *туралы*). Бұл шылаулар ауызекі тілдегі немесе басқа стильдердегі қызметтерімен қатар, ғылыми мәтінде зерттеудің бағытын, нысанын нақтылау, әдістемесі мен құралдарын көрсету, сондай-ақ каузуалды қызмет атқара отырып, академиялық мәтіннің логикалық байланысын, объективтілігін, дәлдігін қамтамасыз етеді. Бұған қатысты мынандай тұрақты қолданыстар анықталды:

- арқылы: *тіл арқылы*, *талдау арқылы*, *әдіс/тәсіл арқылы*, *қолдану арқылы*, *зерттеу арқылы*, *тілдік құралдар арқылы*, *тілдік бірліктер арқылы*, *мысалдар арқылы*, *салыстыру арқылы*, *талдау жасау арқылы*, *тілдік деректер арқылы*, *осы арқылы*, *жүргізу арқылы*, *көрсету арқылы*, *тіркесуі арқылы*, *пайдалану арқылы*, *деректер арқылы*, *жалғау арқылы* т.б.

- үшін: *жасау үшін*, *сілтеме жасау үшін*, *анықтау үшін*, *зерттеу үшін*, *түсіну үшін*, *сипаттау үшін*, *мақсатқа жету үшін*, *талдау үшін*, *жеткізу үшін*, *білдіру үшін*, *көрсету үшін*, *түсіндіру үшін*, *жүргізу үшін*, *қол жеткізу үшін*, *түсінікті болу үшін*, *жүзеге асыру үшін*, *қамтамасыз ету үшін* т.б.

- ретінде: *зерттеу материалы ретінде*, *зерттеу нысаны ретінде*, *қатынас құралы ретінде*, *құбылыс ретінде*, *бірі ретінде*, *жүйе ретінде*, *көрінісі ретінде*, *бірлік ретінде*, *бөлігі*

ретінде, мысал ретінде, символы ретінде, белгісі ретінде, термин ретінде, тұлға ретінде, құрылым ретінде, шет тілі ретінде, екінші тіл ретінде, дереккөз ретінде, ұғым ретінде, категория ретінде, нәтижесі ретінде, концепт ретінде, негізгі материал ретінде, ажырамас бөлігі ретінде т.б.

- байланысты: тығыз байланысты, осыған байланысты, тікелей байланысты, осымен байланысты, өзара байланысты, ерекшеліктеріне байланысты, атауларына байланысты, дамуына байланысты, факторларға байланысты, бір-бірімен байланысты, пайда болуымен байланысты т.б.

- сияқты: сол сияқты, деген сияқты, талдау сияқты, қолдану сияқты, жасау сияқты, бар сияқты, т.б. сияқты, жүйелеу сияқты, айту сияқты т.б.

- қатысты: атауларына қатысты, сөзге қатысты, ұғымына қатысты, концептіге қатысты, тікелей қатысты, оған қатысты, тіліне қатысты, мәселеге қатысты, атауына қатысты, тіл біліміне қатысты, сөз мәдениетіне қатысты т.б.

- туралы: бұл туралы, қаржыландыру туралы, әлем туралы, адам туралы, тіл туралы, мәдениеті туралы, дүние туралы, сипаты туралы, табиғаты туралы, құбылыс туралы т.б.

Академиялық мәтіндерде кездесетін етістіктердің ішіндегі ең жиі қолданылатыны – *де* етістігі. Оның көсемшелі *деп* формасы мен есімшелі *деген* формасының жиілігі жоғары. Бұлар академиялық мәтінде ойды тұжырымдау, пайымдау, сілтеме жасау, цитата келтіру, терминдер мен ұғымдарды анықтау, оларды категориялау қызметтерін атқарады. Сондай-ақ мәтіннің интертекстуалдылығын, объективтілігін арттырып, мәтіннің тұтасымын қамтамасыз етуге қызмет етеді. Біз жинаған корпуста *деп* және *деген* формаларымен тіркесіп келетін мынандай етістікті сөз тіркестері табылды:

- *деп* етістігі: *деп аталатын, деп аталады, деп атайды, деп көрсетеді, деп ойлаймыз, деп атаған, деп санайды/санаймыз, деп есептейді/есептейміз, деп берілген/беріледі, деп жазады, деп тұжырымдайды/тұжырым жасайды, деп түсіндіреді, деп біледі/білеміз, деп сипаттайды/санаймыз, деп таниды, деп жазылады, деп анықтайды, деп түсіндіріледі; деп жауап берген, деп айта аламыз, деп айтуға/атауға болады, деп те атайды, деп анықтама береді, деп атап көрсетеді, деп айта аламыз/айта алмаймыз, деп тануға/санауға/бағалауға болады т.б.* (Сурет 5).

0	деп аталатын	153	24	деп таниды	17	0	деп аталатын синтаксистік	29
1	деп аталады	59	25	деп айтылады	17	1	деп көмекші етістікті	23
2	деп атайды	57	26	деп анықтама	16	2	деп жауап берген	20
3	деп көмекші	57	27	деп беріледі	15	3	деп айта аламыз	17
4	деп көрсетеді	55	28	деп тұжырым	15	4	деп көмекші етістігі	15
5	деп те	55	29	деп жазылады	15	5	деп айтуға болады	15
6	деп атап	52	30	деп анықтайды	14	6	деп атап көрсетеді	13
7	деп ойлаймыз	44	31	деп ойлаймын	13	7	деп көмекші етістігінің	12
8	деп жауап	41	32	деп түсіндіріледі	13	8	деп те атайды	12
9	деп атаған	40	33	деп тануға	12	9	деп анықтама береді	9
10	деп санайды	39	34	деп түсінеміз	12	10	деп атауға болады	9
11	деп есептейді	38	35	деп аталып	12	11	деп тұжырым жасайды	8
12	деп жазады	38	36	деп саналады	11	12	деп аталады бұл	8
13	деп есептейміз	35	37	деп түсінген	11	13	деп көрсетеді бұл	8
14	деп берілген	32	38	деп атауға	11	14	деп айта алмаймыз	7
15	деп санаймыз	28	39	деп береді	11	15	деп тануға болады	7
16	деп тұжырымдайды	26	40	деп айту	10	16	деп санауға болады	7
17	деп айта	25	41	деп атады	10	17	деп бағалауға болады	7
18	деп түсіндіреді	23	42	деп қазақ	10	18	деп жауап берсе	7
19	деп аталған	22	43	деп есептеледі	10	19	деп те аталады	5
20	деп білеміз	20	44	деп танылған	9	20	деп те аталған	5
21	деп көрсеткен	20	45	деп көрсетілген	9	21	деп тұжырым жасауға	5
22	деп айтуға	19	46	деп түр	9	22	деп аталатын желілес	4
23	деп сипаттайды	18				23	деп жазады бұл	4

Сурет 5 – Лингвистикалық академиялық корпустағы «деп» етістігімен тіркесіп келген тіркестер

Рисунок 5 – Словосочетания с глагольной формой «деп»

в Лингвистическом академическом корпусе

Figure 5 – Phrases with the Verbal Form “dep” (saying) in the Linguistic Academic Corpus

- деген етістігі: деген мағына, деген тұжырым, деген пікір, деген анықтама, деген сұрақ, деген көзқарас, деген қорытынды; деген мағынаны білдіреді, деген мағынада қолданылады, деген анықтама берілген, деген пікір айтады, деген пікір білдіреді, деген тұжырым жасайды, деген мағына береді, деген қорытындыға келеді т.б. (Сурет 6).

0	деген мағынаны	77	0	деген мағынаны білдіреді	34
1	деген мағынада	76	1	деген мағынада қолданылған	13
2	деген сөз	74	2	деген мағынада қолданылады	12
3	деген тұжырым	37	3	деген анықтама берілген	11
4	деген пікір	33	4	деген пікір айтады	11
5	деген анықтама	30	5	деген пікір білдіреді	10
6	деген тіркес	26	6	деген мағынаны білдіретін	10
7	деген сөйлемде	24	7	деген тұжырым жасайды	8
8	деген сұраққа	22	8	деген мағынаны береді	8
9	деген сұрақ	21	9	деген мағына береді	8
10	деген сияқты	21	10	деген сұрақтарға жауап	7
11	деген мағына	21	11	деген сұраққа жауап	7
12	деген көзқарасын	20	12	деген мағыналарды береді	7
13	деген мақал	20	13	деген қорытындыға келеді	6
14	деген ұғымды	20	14	деген жыр жолдарында	6
15	деген түсінік	18	15	деген сенім бар	5
16	деген ұғым	16	16	деген түсінік бар	5
17	деген қорытындыға	15	17	деген сөз емес	5
18	деген қорытынды	15	18	деген ұғымды білдіреді	5
19	деген қызығушылық	15	19	деген қаратпа сөздерді	5
20	деген мәселе	15	20	деген қарым қатынасын	4
21	деген тұрақты	14	21	деген қорытынды жасауға	4
22	деген сөйлемдегі	14	22	деген сұрақ туындайды	4
23	деген көзқарасы	14			

Сурет 6 – Лингвистикалық академиялық корпустағы

«деген» етістігімен тіркесіп келген тіркестер

Рисунок 6 – Словосочетания с глагольной формой «деген»

в Лингвистическом академическом корпусе

Figure 6 – Phrases with the Verbal Form “degen” (called) in the Linguistic Academic Corpus

Жоғарыдағы тіркестер академиялық мәтін авторының басқа ғылыми-теориялық еңбектерге сілтеме жасау, шолу жасау арқылы зерттеу тақырыбының маңызын, өзегін және аргументін бекіту үшін, мақсат-пайымын, уәжін көрсету үшін қолданылатын қалыпты құрылымдар екені, бұдан бөлек, зерттеу тақырыбы мен нысанына қатысты терминдер мен ұғымдарды, құбылыстарды атап көрсетуге дайын тұрған тіркестер екені анық байқалады. Бұларды академиялық мәтін сөздер тізімінің маңызды компоненттері, академиялық мәтіннің ажырамас бөлігі деп білеміз.

Академиялық мәтіндерде жиі кездесетін тағы бір белсенді етістік – *бол* көмекші етістігі. Оның *болады* және *болып* формалары жиі қолданылған. Әсіресе, *-уге/уға бол* формантымен келетін тіркестер көп. Бұл – мүмкіндік модальділігін білдіретін формалар. Академиялық мәтін абсолютті шындыққа ұмтылмайды, ол барлық мүмкін сценарийлер мен болжамдарды басшылыққа алып, ықтимал дүниелерді ұсынады. Кейде бұл формалардың белгілі бір ұғымға, құбылысқа, көзқарасқа назар аударту, талдау мен бақылауды қорытындылау үшін де қолданылатын кездері де болады. Бұл формамен келетін мынандай тұрақты қолданыстар анықталды: *байқауға болады, көруге болады, деуге болады, ие болады, негіз болады, пайда болады, атауға болады, анықтауға болады, кездестіруге болады, қарастыруға болады, жасауға болады, айтуға болады, жатқызуға болады, атап өтуге болады, аңғаруға болады, байланысты болады, қолдануға болады, түсіндіруге болады, тануға болады* т.б. (Сурет 7).

0	байқауға болады	134	1	болып саналады	122
1	көруге болады	115	2	болып келеді	121
2	деуге болады	93	3	болып отыр	115
3	ие болады	69	4	болып табылатын	88
4	негіз болады	69	5	болып тұр	37
5	пайда болады	63	6	болып қала	33
6	атауға болады	63	7	болып отырған	27
7	анықтауға болады	47	8	болып саналатын	23
8	кездестіруге болады	45	9	болып есептеледі	23
9	қарастыруға болады	45	10	болып жатқан	21
10	жасауға болады	45	11	болып келетін	19
11	айтуға болады	44	12	болып шығады	14
12	жатқызуға болады	44	13	болып келген	12
13	көрсетуге болады	41	14	болып көрінеді	10
14	өтуге болады	39	15	болып бөлінеді	10
15	байланысты болады	36	16	болып та	9
16	анғаруға болады	34	17	болып қалады	8
17	сөз болады	34	18	болып олардың	8
18	келтіруге болады	33	19	болып шықты	8
19	алуға болады	27	20	болып оның	8
20	пайдалануға болады	26	21	болып тұрған	8
21	қолдануға болады	25	22	болып айтылады	8
22	түсіндіруге болады	24	23	болып қалыптасқан	7
23	беруге болады	24			

Сурет 7 – Лингвистикалық академиялық корпустағы «бол» етістігімен тіркесіп келген тіркестер

Рисунок 7 – Словосочетания с глагольной формой «бол»

в Лингвистическом академическом корпусе

Figure 7 – Phrases with the Verbal Form “bol” (to be/become) in the Linguistic Academic Corpus

Ал *болып* көсемше формалы етістігімен тіркесіп келген мынандай құрылымдар табылды: *болып табылады, болып саналады, болып келеді, болып отыр, болып шығады, болып көрінеді, болып бөлінеді, болып қала береді* т.б.

Академиялық жазылым мәтіндерінде *талдау, зерттеу, көрсету* сынды маркерлік етістіктердің қолданысы жоғары. Бұл етістіктерді автордың ғылыми ой-танымын, ой әрекетін, ғылыми зерттеушілік қабілетін көрсететін академиялық мәтіннің функционалды-семантикалық негізін құрайтын предикатты бірліктер деуге болады. Бұл үш маркерлі етістіктің концептуалды тұрғыдан өзара сабақтасып тұрғанын да байқауға болады. *Талдау* мен *зерттеу* – адами ой-таным қабілетінің ең жоғарғы интеллектуалды мүмкіндігі саналады, ал *көрсету* – талдау мен зерттеуден шыққан нәтижені ұсыну жолы. Сондықтан академиялық мәтіндерде бұл етістіктердің компоненттік байланысы үздіксіз және үнемі деп бағалай аламыз. Мұнда *талдау* мен *зерттеу* сөздерінің заттанған, ұғымдық мәнде қолданысы қатар жүретіндіктен, оның ғылыми мәтіндегі қолданысы тіпті күшейе түседі. Аталмыш етістіктерге қатысты мынандай дайын тілдік құрылымдар табылды:

- Талдау: *талдау арқылы, талдау әдісі/әдістері, талдау барысында/кезінде, талдау жасау, талдау нәтижесінде, талдау негізінде, талдау жүргізілді, талдау және зерттеу, талдау және салыстыру* т.б.

- Зерттеу: *зерттеу барысында, зерттеу нәтижелері, зерттеу нысаны, зерттеу жұмысы, зерттеу нәтижесінде, зерттеу материалы, зерттеу мақсаты, зерттеу әдістері, зерттеу үшін, зерттеуді қажет етеді; зерттеу материалы ретінде, зерттеу нысаны ретінде, зерттеудің ғылыми жаңалығы, зерттеуге мүмкіндік береді, зерттеу жұмысының мақсаты, зерттеу нәтижелері көрсеткендей, зерттеудің ғылыми жаңалығы* т.б.

- Көрсету: *екенін көрсетеді, деп көрсетеді, атап көрсетеді, айқын көрсетеді, маңыздылығын көрсетеді, нәтижелері көрсеткендей, жоғарыда көрсетілген, ашып көрсетеді, сипатын көрсетеді, қажеттілігін көрсетеді, зерттеулер көрсеткендей, ерекшелігін көрсетеді* т.б.

Академиялық мәтіндерде модаль сөздерден *бар, мүмкін, маңызды* бірліктерінің қолданысы жиі болып шықты. Ал *қажет/керек* модаль сөздерінің жиілігі алғашқы елулікке кірмейді: керек сөзінің 750-ге жуық қолданысы табылып 106-орынға, ал қажет сөзі 650-ден аса қолданысымен 124-орынға жайғасқан. Соған қарағанда, қазақ академиялық жазылымы қажеттілік, мүмкіндік,

талап ету модальділігін білдіруге емес, көбіне болжал мәнін білдіруге бейім көрінеді.

Жиналған мәтіндерден аталмыш модаль сөздермен келетін мынандай тіркестер табылды:

- Бар: *мәні бар, маңызы бар, мағынасы бар, ерекшелігі бар, қатысы бар, орны бар, негізі бар, сипаты бар, тікелей қатысы бар* т.б.

- Мүмкін: *болуы мүмкін, мүмкін емес, әкелуі мүмкін, әсер етуі мүмкін, қолданылуы мүмкін, туындауы мүмкін, көрінуі мүмкін, берілуі мүмкін, өзгеруі мүмкін, білдіруі мүмкін* т.б.

- Маңызды: *аса маңызды, өте маңызды, ең маңызды, ерекше маңызды, қарастыру маңызды* т.б.

Академиялық мәтіндерден *тілдік, ұлттық, мәдени, ғылыми, әлеуметтік, негізгі* деген сын есімдердің қолданысы жиі екеніне көз жеткіздік. Бұлар – термин жасаушы, лингвистикаға қатысты ұғымдарды анықтаушы бірліктер. Осы бірліктермен тіркескен мынандай құрылымдар бар: *тілдік бірліктер, тілдік тұлға, тілдік қарым-қатынас, тілдік бейне, тілдік деректер, тілдік құралдар, тілдік жүйе, тілдік сана; ұлттық корпус, ұлттық мінез-құлық, ұлттық сана, ұлттық мәдениет, ұлттық дүниетаным, ұлттық құндылық, ұлттық болмыс, ұлттық таным, ұлттық бірегейлік, ұлттық мәдени код; мәдени код, мәдени ерекшеліктер, мәдени құндылықтар, мәдени құбылыс; ғылыми зерттеу, ғылыми әдістемелік, ғылыми-теориялық, ғылыми еңбектер, ғылыми білімдер, ғылыми жаңалық, ғылыми-практикалық, ғылыми тұрғыдан; әлеуметтік лингвистика, әлеуметтік желі, әлеуметтік мәртебе, әлеуметтік норма, әлеуметтік топ* т.б.

Академиялық жазылымға қажет лексика-грамматикалық минимумды анықтау ғылыми мәтінді меңгерту үшін ғана қажет деп санау қате. Бұл сөздер тобы қазақ тілінің B2-C1 деңгейінде үйренушілер үшін де құнды сөздік дерекқор болмақ. Себебі ғылыми мәтінді кез келген деңгейдегі тіл үйренушілер жаза алмайды. Мәселен, біз ерекше мән бере зерттеген лингво-франка – ағылшын тілінде академиялық сөздік қорды меңгертетін тіл үйренушілерге арналған арнайы оқу құралдары көп. Майкл Маккарти (Michael McCarthy), Фелисити О’Делл (Felicity O’Dell) авторлығымен Cambridge University Press баспасынан жарық көрген “Academic Vocabulary in Use” оқу құралында ғылыми мәтіндерде жиі қолданылатын сөздер мен сөз тіркестерін меңгертуге арналған тапсырмалар, өз бетінше орындайтын жаттығулар берілген. Кітаптың «Академиялық жалпы сөздік қор», «Сөз тіркестері» атты тараушаларында академиялық мәтіндерге тән сөз қолданыстар көрініс тапқан. Авторлар академиялық жалпы сөздік қорға *негізгі зат есімдерді, негізгі етістіктерді, негізгі сын есімдерді, негізгі үстеулерді, академиялық ағылшын тіліне тән фразалық етістіктерді, санды білдіретін негізгі қолданыстарды, көпмағыналы сөздерді, метафоралар мен идиомаларды* жатқызады. Ал сөз тіркестері тараушасында ғылыми мәтінде жиі кездесетін *зат есім және зат есіммен тіркесетін сөздер, зат есім мен сын есім тіркесі, етістік және етістіктен тіркесетін сөздер, септеуліктер мен етістік тіркесі, септеуліктер мен зат есім тіркесі, қысқарған сөздер мен аффикстер* қамтылған. Дәл ағылшын тіліндегідей мұндай жүйелі ғылыми стильді айқындайтын сөздік қор қазақ тілінде нақты анықталып, сұрыпталмаған. Десе де, жоғарыда біз ұсынған тілдік паттерндер, яғни бірнеше бірліктен тұратын тұрақты әрі жиі қолданылатын сөз тіркестері мұндай жүйелі тізімді жасауға қазақ тілінің мүмкіндігі толық жететінін көрсетіп отыр.

«Academic Vocabulary in Use» кітабында жалпы қолданыстағы сөздерден ғылыми мәтінге тән сөздердің бір-біріне өтек ұқсас болатыны, алайда ғылыми мәтінде көбінесе сөздер сәл басқаша мағынада, ауыс мағынада қолданылатыны айтылады. Мысалы, “Underline your family name on the form” деген күнделікті қолданыстан алынған сөйлемде *underline* (астын сызу) тура мағынасында жұмсалса, “The research underlines the value of case studies” сөйлемінде “underline” сөзі «атап көрсету» мағынасын білдіріп тұр (McCarthy, Felicity, 2018: 10).

Қазақ тіл иелмендерінің күнделікті қолданысында жүрген *көрсету, атау* сөздері қазақша академиялық дискурста ауысқан кезде «атап көрсету» тіркесіне айналып шыға келеді. Мәселен: *Бұл зерттеу қолданбалы лингвистикада ЖИ этикалық пайдалануды шаралары бойынша нұсқаулықтардың қажеттілігін көрсетеді және осы саладағы тұтынушылар мен зерттеушілер үшін цифрлық құзыреттердің маңыздылығын атап көрсетеді.* Академиялық стильдің екінші бір белгісі ретінде сөздерге ресмилік тән екені көрсетілген. Мәселен, қазақ

тілінде ауызекі тілде «қысқасы», «қысқасын айтқанда», «тоқетері» десек, ғылыми мәтінде «қорытындылай келгенде», «түйіндейтін болсақ», «қорыта айтар болсақ», «тұжырымдар болсақ» сынды қолданыстарды пайдалану орындырақ саналады.

Қорытынды

Қорытындылай келе айтарымыз, мақалада алғаш рет қазақ тіліндегі тіл білімі саласындағы ғылыми мәтіндердің корпусы жасалып, бұл сандық-корпустық зерттеу қазақша академиялық жазылым мәтіндерінің бастапқы пилоттық нұсқадағы сөз тізімін әзірлеуге мүмкіндік берді. Қазақ академиялық мәтіндерінде басым қолданылатын грамматикалық құралдар мен кейбір лексика-грамматикалық құрылымдар анықталды. Бұл құрылымдар көбіне академиялық мәтіннің тұтасымына, түзіліміне, когезиясына қызмет ететін, сондай-ақ академиялық мәтіннің аргументация, нақтылық, дәлдік, объективтілік сынды қағидаларын, белгілерін сипаттайтын тілдік бірліктер екені байқалды. Зерттеу мақаласының талабына сай, біз жиналған мәтіндер корпусынан ең жиі қолданылатын 50 тілдік бірліктің формаларын, тіркесімдерін ғана қарастырдық. Егер 500 тілдік бірлік немесе 1000 тілдік бірлікке дейінгі жиілікті қарастырсақ, қазақша академиялық жазылым мәтіндерінің басқа да байқала бермейтін құбылыстарын, құрылымдарын анықтауға мүмкіндік болар еді. Мұны келесі зерттеулердің еншісіне қалдырдық. Тағы бір ескеретін жайт, біз тәжірибесін назарға алған ағылшын тіліндегі академиялық сөздер тізімін құрастырудағы негізгі мақсат тіл үйренушілердің тілдік құзыретін арттыру болса, қазақ тіліндегі, соның ішінде тіл білімі саласындағы мақалаларда жиі қолданылатын сөздерді анықтаудағы біздің мақсатымыз – қазақ тілді студенттердің академиялық жазылым дағдысын дамыту, ғылыми аппаратын байытуға көмектесу. Себебі қазіргі таңда жас зерттеушілер мен университет білімгерлеріне академиялық мәтінді жазу сауатының төмендігі, мәтін жазуда ЖИ құралдарын қолданулары, соның салдарынан ЖИ жазған мәтіндегі ұғым-түсініктерді дұрыс түсіндіріп бере алмауына қатысты сын-ескертпелер жиі айтылады.

Біз зерттеуімізде лингвистикалық әдебиеттерде, соның ішінде ғылыми мақалаларда ең жиі көрініс табатын тіл бірліктерін (ең жиі қолданылатын 50 тілдік бірлік пен оның тіркесімін) анықтау арқылы қазақ академиялық жазылымының бір саладағы негізгі лексика-грамматикалық минимумын анықтай алдық. Алайда бұл бағыттағы жұмыс басқа салалардың деректері негізінде де жүзеге асып жалпы қазақ тілді академиялық жазылымға пайдалы тіл бірліктерінің жиынтығын жасау қажет. Себебі барлық саланың академиялық мәтіндеріне ортақ мәтінтүзуші элементтер, лексика-грамматикалық құрылымдар болады. Қазақ тіліндегі дәл осы құрылымдардың алғашқы нұсқалары анықталды деуге негіз бар.

Зерттеу мынандай нәтижелерге қол жеткізді:

- Тіл білімі бағытындағы қазақ академиялық мәтіндерінде жиі қолданылатын лексика-грамматикалық құралдар анықталды;
- Тіл білімі бағытындағы қазақ академиялық мәтіндерін тұтастырушы, түзуші, оның когезиясын қамтамасыз етуші лексика-грамматикалық құрылымдар сипатталды;
- Тіл білімі бағытындағы қазақ академиялық мәтіндерінде ең жиі қолданылатын тілдік бірліктер сұрыпталды;
- Тіл білімі бағытындағы қазақ академиялық мәтіндеріне тән кей лексика-грамматикалық, концептуалдық ерекшеліктер көрсетілді.

Зерттеудің негізгі түйіні және бұл тізімді академиялық жазылым курстарын жетілдіруге, қазақ академиялық сөздерінің сөздігін жасауға, NLP (табиғи тілді өңдеу) тапсырмаларын орындауға, қазақ тілін шет тілі ретінде оқыту тәжірибесінде және ғылыми дискурс бағытындағы басқа да зерттеулерге қолдануға мүмкіндік бар.

Әдебиеттер

Қазақ тілінің ұлттық корпусы [Электронды ресурс]. – Ахмет Байтұрсынұлы атындағы Тіл білімі институты. – URL: <https://qazcorpus.kz/> (қаралған күні: 24.06.2026).

Қоңыратбаева Ж.М. Қазақ тілінің академиялық корпусын әзірлеу: мақсаты, міндеті, маңызы (гуманитарлық ғылымдар бағытында). – Қарағанды университетінің хабаршысы. – Филология сериясы, 2025, 30, 3(119). – 88-98-бб. <https://doi.org/10.31489/2025Ph3/88-98>

Негізгі корпус мәтіндерінің жиілік сөздігі. – Алматы: Асыл ЖК, 2025. – 1432 б.

Сарсеке Г., Акмагамбетова Б., Сентірбек Ж. Академиялық корпус және оны қазақ тілінде әзірлеудің өзектілігі мен маңыздылығы. – Вестник Кокшетауского университета имени Шокана Уалиханова. Филологическая серия. – 2025. – №2. – С. 213–227. <https://doi.org/10.59102/kufil/2025/iss2pp213-227>

Coxhead, A. An academic word list. School of Linguistics and Applied Language Studies, Victoria University of Wellington, 1998. Vol. 18. https://www.academia.edu/download/33317539/LDOCE_AWL.pdf [Электронды ресурс] (қаралған күні: 24.06.2026).

Coxhead, A. A New Academic Word List. TESOL Quarterly. 2000, 34: 213-238. <https://doi.org/10.2307/3587951>

Durrant, Ph. To what extent is the Academic Vocabulary List relevant to university student writing? English for Specific Purposes. Vol. 43. 2016. Pp. 49-61. <https://doi.org/10.1016/j.esp.2016.01.004>.

McCarthy, M., Felicity, O. Academic Vocabulary in Use. Cambridge: Cambridge University Press, 2018. 175 p.

Odanova, S.A., Tuleup, M.M., Moldabayeva, K.E. The Origin Of Academic Writing Style In Kazakh Linguistics. Journal “Bulletin of the Kazakh State University named after Abylai Khan”, Series “Philological Sciences”. Vol. 74. No. 3 (2024). <https://doi.org/10.48371/PHILS.2024.3.74.011>

Öner, M., Moldabayeva, K., Odanova, S. Basics For Designing A Digital Simulator For Academic Writing Skills Formation. *Tiltanyim*. 2024;(3): 151-162. <https://doi.org/10.55491/2411-6076-2024-3-151-162>

Paquot, M. Academic Vocabulary in Learner Writing: From Extraction to Analysis. London; New York: Continuum, 2010. 266 p.

References

Coxhead, A. (1998) An academic word list. School of Linguistics and Applied Language Studies, Victoria University of Wellington, Vol. 18. https://www.academia.edu/download/33317539/LDOCE_AWL.pdf [Electronic resource] (accessed: 24 June 2026). (in English)

Coxhead, A. (2000) A New Academic Word List. TESOL Quarterly. 34: 213-238. <https://doi.org/10.2307/3587951> (in English)

Durrant, Ph. (2016) To what extent is the Academic Vocabulary List relevant to university student writing? English for Specific Purposes. Vol. 43. Pp. 49-61. <https://doi.org/10.1016/j.esp.2016.01.004>. (in English)

McCarthy, M., Felicity, O. (2018) Academic Vocabulary in Use. Cambridge: Cambridge University Press. 175 p. (in English)

National Corpus of the Kazakh Language [Electronic resource] (2026). Ahmet Baitursynuly Institute of Linguistics. URL: <https://qazcorpus.kz/> (accessed: 24 June 2026). (in Kazakh)

Negizgi korpus matinderining zhiilik sozdigi (2025) Almaty: Asyl ZhK. 1432 b. [Frequency Dictionary of main corpus texts. Almaty: Asyl ZhK. 1432 p.] (in Kazakh)

Odanova, S.A., Tuleup, M.M., Moldabayeva, K.E. (2024) The Origin Of Academic Writing Style In Kazakh Linguistics. Journal “Bulletin of the Kazakh State University named after Abylai Khan”, Series “Philological Sciences”. Vol. 74. No. 3 <https://doi.org/10.48371/PHILS.2024.3.74.011> (in English)

Öner, M., Moldabayeva, K., Odanova, S. (2024) Basics For Designing A Digital Simulator For Academic Writing Skills Formation. *Tiltanyim*. No. 3: 151-162. <https://doi.org/10.55491/2411-6076-2024-3-151-162> (in English)

Paquot, M. (2010) Academic Vocabulary in Learner Writing: From Extraction to Analysis. London; New York: Continuum. 266 p. (in English)

Qongyratbaeva, Zh.M. (2025) Qazaq tilining akademijalyq korpusyn azirleu: maqsaty, mindeti, mangyzy (gumanitarlyq gylymdar bagytynda). Qaragandy universitetining Habarshysy. Filologiya serijasy, 30, 3(119). 88-98-bb. [Konyratbayeva, Zh.M. (2025) Development of the academic Corps of the Kazakh language: purpose, objectives, significance (in the direction of Humanities). Bulletin of Karaganda University. Philology series, 30, 3(119). Pp. 88-98.] <https://doi.org/10.31489/2025Ph3/88-98> (in Kazakh)

Sarseke, G., Akmagambetova, B., Sentirbek, Zh. (2025) Akademijalyq korpus zhane ony qazaq tilinde azirleuding ozektiligi men mangzydylygy. Sh. Ualihanov atyndagy Kokshetau Universitetining Habarshysy. Filologiya serijasy. No. 2. 213–227-bb. [Sarseke, G., Akmagambetova, B., Sentyrbek, Zh. (2025) Academic corpus and the relevance and importance of its development in the Kazakh language. Bulletin of Sh. Ualikhanov Kokshetau University Philological Series. No. 2. Pp. 213-227.] <https://doi.org/10.59102/kufil/2025/iss2pp213-227> (in Kazakh)

Мақала туралы ақпарат / Информация о статье / Information about the article

Редакцияға түсті / Поступила в редакцию / Entered the editorial office: 03.07.2026.

Жариялауға қабылданды / Принята к публикации / Accepted for publication: 23.09.2026.

Интернетте қолжетімді / Доступно онлайн / Available online: 30.09.2026.