

С.Қ. Құлманов^{1*}, А.З. Бисенғали¹, Н.Қ. Шүленбаев¹, А.Д. Омарова²

¹А. Байтұрсынұлы атындағы Тіл білімі институты, Қазақстан, Алматы қ.

²Әл-Фараби атындағы Қазақ ұлттық университеті, Қазақстан, Алматы қ.

*e-mail: k.k.sarsen@mail.ru

ҚАЗАҚ ЦИФРЛЫҚ ТЕРМИНОЛОГИЯСЫ: ТЕРМИНОЛОГИЯЛЫҚ ІШКОРПУС ӘЗІРЛЕУ ТӘЖІРИБЕСІ (метамәтіндік және терминологиялық белгіленімдер параметрлері)

Аннотация. Мақала Қазақ тілі ұлттық корпусының (ҚТҰК) құрамына енетін терминологиялық ішкорпус жасаудың ғылыми-теориялық және қолданбалы мәселелеріне (метамәтіндік және терминологиялық белгіленімдер параметрлерін әзірлеу) арналды.

Зерттеу барысында ҚТҰК-ның терминологиялық ішкорпусын әзірлеу бойынша мынадай нәтижелерге қол жеткізілді: 1) терминологиялық ішкорпус базасына енгізілген ғылыми мәтіндерге төмендегідей метамәтіндік белгіленімдер қойылды: Дереккөзі, Мәтін типі, Мәтін тақырыбы, Мәтін стилі, Мәтіннің кілт сөздері, Ғылым саласы, Мәтін тілі, Мәтіннің бет саны, Мәтін графикасы, Сөзқолданыс саны, Аудитория жасы, Субкорпус, Мәтіннің енгізілген уақыты, Белгілеуші; 2) терминологиялық ішкорпусқа енгізілген әртүрлі ғылым салаларының терминдеріне мынадай параметрлерді қамтитын терминологиялық белгіленімдер қойылды: терминнің қазақ және орыс тілдеріндегі нұсқасы, терминнің және терминдік тіркесті құрайтын сөздердің этимологиясы және оның қазақ тіліндегі баламасы, әрбір терминнің қазақ тіліндегі анықтамасы, терминнің анықтамасы алынған дереккөз, терминнің қолданылатын саласы (салалары), терминнің авторы, терминнің құрамы, терминнің бекітілген нұсқасы, жылы, терминнің салалық терминологиялық сөздіктердегі аудармасы, терминнің заңнамадағы қолданысы (аудармасы және нұсқалары).

Зерттеудің теориялық маңыздылығы: екітілді және көптілді салалық, түсіндірме терминологиялық сөздіктер құрастырудың, электронды сөздіктер әзірлеудің ғылыми негіздерін нақтылап, толықтыруға, қолданбалы терминология, терминография салаларын дамытуға септігін тигізеді.

Зерттеудің практикалық маңыздылығы: тақырыптық және салалық терминологиялық ішкорпустар әзірлеу мен әртүрлі терминологиялық сөздіктердің (түсіндірме, реестрлік) құрылымы мен мазмұнын жетілдіру барысында лексикографиялық тәжірибеде қолдануға болады.

Зерттеу барысында жинақтау, сұрыптау, жүйелеу, аудару, диахрондық, фактологиялық, лексикографиялық салыстыру және салғастыру, құрылымдық (морфологиялық, синтаксистік) талдау әдістері қолданылды.

Тірек сөздер: терминология, ішкорпус, метамәтіндік белгіленім, терминологиялық белгіленім, сөздік.

С.К. Кулманов^{1*}, А.З. Бисенғали¹, Н.Қ. Шүленбаев¹, А.Д. Омарова²

¹Институт языкознания имени А. Байтұрсынова, Казахстан, г. Алматы

²Казахский национальный университет имени аль-Фараби, Казахстан, г. Алматы

*e-mail: k.k.sarsen@mail.ru

КАЗАХСКАЯ ЦИФРОВАЯ ТЕРМИНОЛОГИЯ: ОПЫТ РАЗРАБОТКИ ТЕРМИНОЛОГИЧЕСКОГО ПОДКОРПУСА (параметры метатекстовых и терминологических разметок)

Аннотация. Статья посвящена научно-теоретическим и прикладным вопросам (разработка метатекстовых и терминологических разметок) создания терминологического подкорпуса, входящего в состав Национального корпуса казахского языка (НККЯ).

В ходе исследования были достигнуты следующие результаты по разработке терминологического подкорпуса НККЯ: 1) научные тексты, входящие в терминологическую подкорпусную базу, имеют следующие метатекстовые разметки: Источник, Тип текста, Заголовок текста, Стили текста, Ключевые слова текста, Область науки, Язык текста, Количество страниц текста, Графика текста, Количество словоупотреблений, Возраст аудитории, Субкорпус, Дата внесения текста, Разметчик; 2) терминологическим терминам различных отраслей науки были присвоены терминологические разметки, включающие следующие параметры: варианты термина на казахском и русском языках, этимология термина и слов, составляющих терминологическое сочетание, и его эквивалент в казахском языке, определение каждого термина на казахском языке, источник, из которого было получено определение термина, область (области) применения термина, автор термина, состав термина, утвержденный вариант термина, год его утверждения, перевод термина в отраслевых терминологических словарях, использование термина в законодательстве (перевод и варианты).

Теоретическая значимость исследования: разъясняет и дополняет научные основы составления двуязычных и

многоязычных отраслевых, толковых терминологических словарей, разработки электронных словарей, способствует развитию прикладной терминологии и терминографии.

Практическая значимость исследования: может быть использована в лексикографической практике в процессе разработки тематических и отраслевых терминологических подкорпусов и во время совершенствования структуры и содержания различных терминологических словарей (толковых, реестровых).

В ходе исследования были использованы методы сбора, сортировки, систематизации, перевода, диахронного, фактологического, лексикографического сравнения и сопоставления, структурного (морфологического, синтаксического) анализа.

Ключевые слова: терминология, подкорпус, метатекстовая разметка, терминологическая разметка, словарь.

S.K. Kulmanov^{1*}, A.Z. Bissengali¹, N.K. Shulenbayev¹, A.D. Omarova²

¹A. Baitursynuly Institute of Linguistics, Kazakhstan, Almaty

²Al-Farabi Kazakh National University, Kazakhstan, Almaty

*e-mail: k.k.sarsen@mail.ru

KAZAKH DIGITAL TERMINOLOGY: EXPERIENCE IN DEVELOPING A TERMINOLOGICAL SUBCORPUS (metatext and terminology markup parameters)

Annotation. The article is devoted to scientific-theoretical and applied-practical issues (development of metatext and terminological markings) of the creation of a terminological subcorpus, which is part of the National Corpus of the Kazakh Language (NCKL).

In the course of the study, the following results were achieved in the development of the terminological subcorpus of the NCKL: 1) scientific texts included in the terminological corpus base have the following meta-text markups: Source, Text type, Text Title, Text Style, Keywords of the text, Field of science, Text language, Number of pages of text, Text graphics, Number of usages, Age of the audience, Subcorpus, Date the text was entered, Markup; 2) terminological terms of various branches of science were assigned terminological markings, including the following parameters: variants of the term in Kazakh and Russian, the etymology of the term and the words that make up the terminological combination, and its equivalent in the Kazakh language, the definition of each term in the Kazakh language, the source from which it was obtained definition of the term, area(s) of application of the term, author of the term, composition of the term, approved version of the term, year of its approval, translation of the term in industry terminology dictionaries, use of the term in legislation (translation and variants).

The theoretical significance: clarifies and complements the scientific foundations of the compilation of bilingual and multilingual industry, explanatory terminological dictionaries, the development of electronic dictionaries, contributes to the development of applied terminology and terminography.

The practical significance: it can be used in lexicographic practice in the process of developing thematic and sectoral terminological subcorps and during the improvement of the structure and content of various terminological dictionaries.

The research used methods of collection, sorting, systematization, translation, diachronic, factual, lexicographic comparison and matching, structural (morphological, syntactic) analysis.

Keywords: terminology, subcorpus, metatextual markup, terminological markup, dictionary.

Кіріспе

Кез келген тілдің даму тарихы тілдік бірліктердің грамматикалық құрылымындағы және семантикалық мазмұнындағы өзгерістермен, тілдің қоғамдық-әлеуметтік қызметімен байланысты болғандықтан, ғылыми лингвистикалық парадигмада қолданылатын зерттеу тәсілдері мен аспектілері де өзгеріп, жетілдіріліп отырады. Бұл қазіргі тіл білімінде мәтінді зерттеу объектісі ретінде қарастыруда материалды өңдеуді оңтайландыру мен оңайлату мәселесін күрделендіре түсті. Осыған байланысты ақпараттық технологияның қарыштап жетілуіне орай тілдік практикалық материалды жинау мен талдау мұқият әзірленген қағидаттар мен механизмдерді талап етіп, тіл білімі мен IT-технология саласы тоғысында пайда болған «корпустық лингвистика» ғылымы бүгінгі күні қарқынды дамып келеді.

Әлем тілдеріндегі қолданбалы (корпустық, компьютерлік) тіл білімінің жетістіктерін лингвистикалық ұлттық корпустарды зерттеу мен әзірлеу мақсатында пайдалану тіл білімі және IT-технология саласы үшін ғана емес, жалпы әлеуметтік-қоғамдық мәселе саналады. Сонымен қатар лингвистикалық корпустарды жасаудан алынған нәтижелерді іске асыру, атап айтқанда, оларды әртүрлі сөздіктер құрастыруда, әртүрлі (академиялық, нормативтік) грамматикалар жазуда немесе белгілі бір лингвистикалық құбылыстарға сипаттама беруде, мәтіндік материалды автоматты жолмен өңдеуде орасан пайдасын тигізіп қана қоймай, ғылым салаларында зерттеулер

жүргізуде жаңа технологиялар қалыптастыруға тиімді жол ашады. Әсіресе пән салалары терминдерінің күннен күнге көбейіп, қарқынды қолданылуы қолданбалы тілтаным саласында ғылыми мәтіндерді, техникалық құжаттаманы және т.б. автоматты түрде өңдеу бағытындағы зерттеулердің өзектілігін күшейте түсті. Осы ретте бұл мақала қазақ терминологиялық ішкорпусын жасаудың ғылыми-теориялық және қолданбалы-практикалық мәселелеріне (метамәтіндік және терминологиялық белгіленімдер параметрлерін әзірлеу) арналып отыр.

Материал және әдістер

Терминологиялық ішкорпус қазақ тілі терминдерін жүйелеу мен біріздендіру, одан әрі қалыптандыру (кодификациялау, кодтау) мен стандарттауға арналған жүйелі электронды (цифрлық) қор болғандықтан, оны әзірлеу барысында салалық терминдерді *жинақтау*, *сұрыптау*, *жүйелеу* тәрізді жалпылингвистикалық әдістер, терминдердің дефиницияларын (анықтамаларын) *аудару*, әртүрлі дереккөздердегі анықтамаларымен *салыстыру* әдісі, терминнің және терминдік тіркесті құрайтын бірліктердің этимологиясын табуда *диахрондық* әдіс, терминнің анықтамасы алынған дереккөзді, терминнің авторын, терминнің бекітілген нұсқасы мен жылын, терминнің салалық терминологиялық сөздіктердегі аудармасы мен терминнің заңнамадағы қолданысын айқындауда *фактологиялық*, *лексикографиялық салыстыру* және *салғастыру* әдістері, терминнің құрамын ажыратуда *құрылымдық (морфологиялық, синтаксистік) талдау* әдістері қолданылды.

Лингвистикалық корпустар жасаудың әлемдік тәжірибесін зерделеу барысында терминологиялық ішкорпус әзірлеудің негізгі әдіс-тәсілдері, өлшемшарттары (критерийлері), терминологиялық белгіленімдердің түрлері, тілдік және тілден тыс ақпаратты беру, әртүрлі корпустардағы терминологиялық ақпаратты өңдеу (енгізу, алу және т.б.), терминдерді аудару, терминографиялық лексикографияны жетілдіру, терминологиялық жұмыстарға электрондық мәтіндік корпустарды пайдалану және т.б. мәселелер қарастырылған ғылыми еңбектер басшылыққа алынды. Сонымен қатар терминологиялық ішкорпус базасына қажетті материалдарды жинақтау үшін және қазақ тілі терминологиялық қорының эволюциясын және қазіргі жай-күйін зерделеу мақсатында әртүрлі ғылым салаларына арналған сөздіктер пайдаланылды.

Әдебиетке шолу

Корпустық лингвистика қазіргі кезде қолданбалы (корпустық, компьютерлік) тіл білімі мен IT-технология саласының ұштасқан пәнаралық ғылым ретінде жалпы тіл білімінде қарқынды дамып отыр. Осы ретте Қазақстанда да тіл білімін цифрландыру бойынша көптеген ғылыми зерттеулер жүргізіліп, әртүрлі қолданбалы құралдар әзірленіп жатқандығын атап өтеміз.

Қазақ тіл біліміндегі цифрландыру бойынша жұмыстарды төмендегідей бағыттарға бөлуге болады:

– параллель корпустар әзірлеу (Аязбаев, 2023: 49-61);

– әртүрлі цифрлық лингвистикалық құралдар: Atau (<https://atau.kz/kz/>) – ономастикалық атауларға арналған платформа; Emle.kz (<https://emle.kz/>) – емле ережелеріне, сөздердің дұрыс жазылуына арналған цифрлық платформа; Qazlatyn (<https://www.qazlatyn.kz/>) – қазақ тіліндегі мәтіндерді кирил қарпінен латын әрпіне және керісінше «аударатын» құрал; Qijat (<https://qijat.kz/>) – ресми құжаттарға қол жеткізуге, оларды жолдауға немесе өңдеуге арналған платформа; Sozdikqor.kz (<https://sozdikqor.kz/>) – қазақ тіліндегі сөздер мен сөз тіркестері базасынан тұратын онлайн-сөздік; Termincom (<https://termincom.kz/>) – қазақ-орыс тілдеріндегі салалық терминологиялық қорға арналған терминологиялық дерекқор; Til-Qural (<https://tilqural.kz/>), TilMedia (<https://tilmedia.kz/>) – қазақ тілін оқытуға арналған тегін онлайн платформа (Баекеуева, 2022: 83-98);

– «Алматы қазақ тілінің корпусы» (<http://web-corpora.net/KazakhCorpus/search/>) – зерттеушілер мен тілдік мәтінді өңдеу саласындағы әзірлемешілерге арналған қазақ тіліндегі мәтіндер жиыны;

– «Қазақ тілінің ұлттық корпусы» (КТҰК) (<https://best.qazcorporaz.kz/eng/>) – лингвистикалық зерттеулер мен әзірлемелерге арналған қазақ тіліндегі мәтіндер қоры.

Қазіргі тілдік жүйеде адам баласының ойлау қызметінің вербалданған нәтижесінің бірі

ретіндегі терминдер қоғамның барлық саласында қолданылады. Өйткені, терминдер, біріншіден, тарихи тұрғыда ұлттық танымдық ойдың бір бөлшегін құрайды, яғни қандай да бір саланың белгілі бір даму кезеңіндегі білімді көрсетеді; екіншіден, олар белгілі бір пән саласы ғылыми тілінің дамуына септігін тигізеді; үшіншіден, арнаулы мақсатта қолданылғанда тілдің шеңберінен «шығып кетіп», белгілі бір салада жинақталған немесе «сыртқы түрткіжайттардың» әсерінен қалыптасқан білім мен тәжірибені сақтау мен таратуға арналған ұғымдық бірлік болып табылады (Бутенко, 2021: 47-48).

Ғылым мен техниканың дамуына байланысты жалпы лингвистикада тілдік құбылыстарды, оның ішінде терминологияны корпустық лингвистика саласының әдіс-тәсілдері, әртүрлі технологиялық құралдар арқылы зерттеу кеңінен қолданылып келеді. Осы бағыттағы негізгі зерттеулерге мыналарды жатқызуға болады: корпустардан жасанды интеллект саласына қатысты терминологияны жартылай автоматты түрде алуға арналған ізденістер (Mikelionienė, 2021: 30-38); БҰҰ корпусын прагматикалық, терминологиялық және функционалды талдау бойынша ғылыми зерттеулер (Balashova, 2022: 25-43); мектеп оқулықтарында қолданылатын терминологияны корпус арқылы зерттеу (Monakhov, 2023: 27-49); корпусқа академиялық мәтіннің лексикалық параметрлерін енгізуге арналған бағдарлама әзірлеу (Martynova, 2020: 72-80); ғылыми және техникалық мәтіндер корпусын әзірлеудің концептуалды негізін анықтауға арналған ізденістер (Butenko, 2021: 76-81); интеллектуалды ғылыми және білім беру интернет-ресурстарын (ISEIR) жасау мен сүйемелдеуге арналған әдістерді, технологиялар мен құралдарды әзірлеу бойынша зерттеулер; табиғи тілді өңдеудің қазіргі заманғы әдістеріне негізделген интеллектуалды ақпараттық ресурс моделін (IPR NLP) әзірлеудің ғылыми негіздерін айқындау (Sadirmekova, 2023: 5314-5332); тілдік коммуникацияны өңдеу және сапасын арттыру мақсатындағы жазылым мен аудармаға арналған цифрлық құралдар әзірлеу (Tazhibayeva, 2023: 166-174) және т.б.

Қазақ тілі білімінде корпустық лингвистика саласы, оның ішінде ұлттық корпус жасау, әртүрлі ішкорпустар әзірлеу бойынша жұмыстар 2000-жылдардың басында қолға алынып, қазіргі кезде айтарлықтай нәтижелерге жетіп отыр. Осы ретте қазақ тілі білімінде корпустық лингвистика саласының әрі Қазақ тілінің ұлттық корпусын әзірлеудің ғылыми-практикалық негізін салушы филология ғылымдарының докторы, профессор А.Қ. Жұбановтың есімін ерекше атап өтеміз. Сондай-ақ Қазақ тілінің ұлттық корпусын әзірлеудің тұжырымдасын жасаған филология ғылымдарының докторы, профессор Э.Д. Сүлейменованың, Қазақ тілінің ұлттық корпусының қазіргі деңгейіне жетуіне ғылыми-ұйымдастырушылық басшылық жасап келе жатқан филология ғылымдарының кандидаты А.М. Фазылжанованың, ҚТҰҚ және оның ішкорпустарын жасауға ғылыми, үйлестірушілік жетекшелік етіп келе жатқан филология ғылымдарының докторы А.Ә. Жаңабекованың еңбектері орасан зор. Осындай мықты ғылыми-практикалық негізі жасалған Қазақ тілінің ұлттық корпусы бүгінгі күні жан-жақты зерттеліп, әртүрлі ішкорпустар жасау ісі жолға қойылған. Мысалы, фонетикалық және просодикалық маркерлерді әзірлеу (Bazarbayeva, 2023: 1256-1270), ономастикалық ішкорпус жасау және мифонимдер мен теонимдердің репрезентативтілігін анықтау мәселесі зерделенеді (Tlegenova, 2024: 374-395). Сонымен қатар жасалып жатқан ішкорпустарға белгіленім қою күрделі мәселе саналады. Осы ретте төмендегі ғылыми зерттеулер мынадай тақырыптарға арналған: тарихи ішкорпус құрамына енгізілген мәтіндерге белгіленім қою (Сейітбекова, 2022: 159-170); диалектілік ішкорпус әзірлеудің технологиясы (Zhanabekova, 2022: 70-87); этномәдени ішкорпус жасау тәжірибесі (Исаева, 2023: 135-143); қазақ және орыс тілдеріндегі көркем әдебиет пен заңнамалық құжаттардың параллель ішкорпусын әзірлеу (Аязбаев, 2023: 49-61); әртүрлі сипаттағы жарнамалардың ішкорпусын әзірлеу (Күркімбаева, 2023: 164-173) және т.б.

Нәтижелер және талқылау

Мақаланың негізгі нысаны қазақ тілінің терминологиялық ішкорпус мәселесі болғандықтан, Ахмет Байтұрсынұлы атындағы Тіл білімі институтында әзірленіп жатқан ҚТҰҚ негізгі корпусына және оның ішкорпустарына кеңірек тоқталуды мақсат еттік. *Негізгі корпус* – көркем әдебиет, ғылыми, публицистикалық, ісқағаз, сөйлеу тілі стилі мәтіндерінің ғылыми зерттеу және оқыту ІТ-ресурсы ретіндегі электронды жиынтығы. Негізгі корпус сөз, сөзформа

(сөз түрленімі) бойынша іздеу жүйесінен және морфологиялық, сөзжасамдық, семантикалық, лексикалық, фонетика-фонологиялық белгіленім түрлерінен тұрады. Корпусқа енгізілген мәтіндерге 10 параметрден тұратын метамәтіндік белгіленім қойылған: *автор, мәтін атауы, мерзімі, дереккөз, стиль, таралым типі, аудитория жасы, жынысы*. Негізгі корпусның мәтіндер базасы төмендегідей: көркем әдебиет – 7167348; поэзия – 1000000; оқулық мәтіндері – 2000000; публицистикалық – 11950995; ғылыми – 7779000; іскери – 2208448; сөйлеу – 1000000). *Ахмет Байтұрсынұлы ішкорпусы* – Ахмет Байтұрсынұлының өлеңдері мен әңгімелері, мақалалары, оқу құралдары мәтіндерінің электронды дерекқоры (132 000 сөзқолданыс); *Ауызша ішкорпус* – аудио/бейнежазбалар транскрибацияларының электронды дерекқоры (1000100 сөзқолданыс); *Диалектілік ішкорпус* – жергілікті тіл ерекшеліктерін қамтыған мәтіндердің электронды жиынтығы (3 484 440 сөзқолданыс); *Лингвомәдени ішкорпус* – мәдени-танымдық мағынасы жасырын тұрған тілдік бірліктердің астарындағы білімдерді анықтауға негізделген электрондық база (157 әдебиет); *Мақал-мәтелдер ішкорпусы* – уәждік түсіндірмесі берілген мақал-мәтелдерден тұратын электрондық база (138 647 сөзқолданыс); *Тарихи ішкорпус* – ескі кезеңдерден бері жарық көрген әртүрлі графикалық жүйедегі жазба мұраларға метамәтіндік белгіленімдер қойылған электрондық база (1000200 сөзқолданыс); *Ономастикалық ішкорпус* – ономастикалық мәліметтер қамтылған электронды ақпараттық-анықтамалық жүйе (1000000 сөзқолданыс); *Жарнама ішкорпусы* – таратылу орны мен әдісіне қарай жіктелген жарнама мәтіндерінің жиынтығы (4000 бірлік); *Параллель корпус* – түпнұсқа көркем әдебиет және заңнама мәтіндері және олардың аудармаларының жиынтығы (2100000 сөзқолданыс); *Терминологиялық ішкорпус* – әртүрлі салаларға арналған ғылыми еңбектерден жинақталып, жүйеленген, метамәтіндік белгіленімдер қойылған және терминдерге (8000 термин) терминологиялық белгіленім қойылған мәтіндер жиынтығы (3300000 сөзқолданыс).

Қазақ тілінің ұлттық корпусының терминологиялық ішкорпусын әзірлеу төмендегідей кезеңдер мен сатылардан тұрды:

I ҚТҰК-ның терминологиялық ішкорпусына қажет тілдік материалдарды сұрыптау, жүйелеу, өңдеу

ҚТҰК-ның терминологиялық ішкорпусына әртүрлі ғылым салаларына арналған еңбектерден жиналып, сұрыпталып, жүйеленген және техникалық өңдеуден өткізілген, 3 300 000 сөзқолданыстан тұратын ғылыми мәтін енгізілді.

ҚТҰК-ның терминологиялық ішкорпусының базасына енгізілген ғылыми мәтіндердің әрқайсысына төмендегідей метамәтіндік белгіленімдер қойылды:

1) Дереккөзі. Ғылыми мәтін алынған еңбектің толық деректері көрсетілді. Мысалы, Какимов У.К., Каипова А.А., Елемесов К.К., Басқанбаева Д. *Металлургия цехтарының газ тазарту және шаңжинағыш жабдықтары*. – Алматы, 2010. – 245 б.

Мұнда егер мәтін интернет көзінен алынған болса, оның сілтемесі көрсетілді.

2) Мәтін типі. Бұл метамәтіндік белгіленімге *монография, оқулық, оқу құралы, мақала, тезис, реферат, аңдатпа, түйіндеме, баяндама, пікір* және т.б. сияқты түрлер енгізіліп, аталған метамәтіндік белгіленім ішінен ғылыми мәтіннің сипатына қарай оның типі таңдалды. Мысалы, *монография* және т.б.

3) Мәтін тақырыбы. Мұнда ғылыми мәтіннің қандай ғылыми мәселеге (тақырыпқа) арналғаны көрсетілді. Мысалы, *металлургия цехтарының газ тазарту және шаңжинағыш жабдықтары*.

4) Мәтін стилі. Бұл метамәтіндік белгіленім ғылыми мәтіннің стилін көрсетуге арналды. Мысалы, мәтіннің сипатына байланысты *ғылыми, ғылыми-көпшілік, ғылыми-танымдық, ағартушылық, аралас* деген түрлердің бірі таңдалды.

5) Мәтіннің кілт сөздері. Терминологиялық ішкорпустан терминдерді іздеу оңай әрі тез жүру үшін ғылыми мәтінді сипаттайтын 5 сөзден аспайтын кілт сөз (немесе сөз тіркесі) көрсетілді. Мысалы, *агломерациялық машина, мартен газы, электр доғалық пеш, газды тазарту, газ сору* және т.б.

6) Ғылым саласы. Бұл метамәтіндік белгіленімде ғылыми мәтіннің саласы көрсетілді. Мысалы, *металлургия* және т.б.

7) Мәтін тілі. Терминологиялық ішкорпусқа қазіргі кезде қазақ тіліндегі ғылыми мәтіндер енгізіліп жатқандықтан, мұнда «мәтін тілі» деген бағанға «қазақ» деп көрсетілді. Сонымен қатар болашақта басқа тілдерді қамту және параллель корпуспен байланыстыру көзделеді.

8) Мәтіннің бет саны. Мысалы, 500.

9) Мәтін графикасы. Терминологиялық ішкорпусқа қазіргі кезде кирил қарпімен жазылған ғылыми мәтіндер енгізіліп жатқандықтан, бұл бағанға «кирил» деп көрсетілді. Сондай-ақ терминологиялық ішкорпусқа болашақта латын және араб графикасымен жазылған ғылыми мәтіндерді енгізу көзделетіндіктен, «латын», «араб» деген белгіленім түрлері қойылды.

10) Сөзқолданыс саны. Мысалы, 58 000.

11) Аудитория жасы. Терминологиялық ішкорпусқа енгізілетін ғылыми мәтіндер негізінен студенттерге, магистранттар мен докторанттарға, ғалымдарға арналғандықтан, мұнда «ересек» деп көрсетілді, сондай-ақ мектеп оқулықтары, газет-журналдардағы ғылыми-көпшілік, ғылыми-танымдық мәтіндер де енгізілетіндіктен және әртүрлі жастағы тұтынушыларға арналғандықтан, мұнда «бала», «бейтарап» деген белгіленім түрлері қойылды.

12) Субкорпус. Қазіргі әзірленіп жатқан терминологиялық ішкорпустың болашақта «ғылыми-техникалық» деп аталатын субкорпусын жасау көзделетіндіктен, осы белгіленім түрі көрсетілді.

13) Мәтіннің енгізілген уақыты. Мысалы, 15.11.2023.

14) Белгілеуші. Мұнда терминологиялық ішкорпусты жасауға қатысушы, яғни осы мәтінді енгізген қызметкердің аты-жөні көрсетілді. Мысалы, А.Ж. Үдербаев.

1-сурет – ҚТҰК-ның терминологиялық ішкорпусының базасына енгізілген ғылыми мәтінге қойылған метамәтіндік белгіленім үлгісі

II Терминологиялық ішкорпустың терминологиялық белгіленім жүйесін жасау

Терминологиялық ішкорпустың терминологиялық белгіленім жүйесі мынадай параметрлерден тұрады:

1) төл және шет тілдерінен енген термин сөз немесе терминдік тіркес: мысалы, *сөзжасам*, *еңбек бөлінісі*, *абдукция*, *дисконт* және т.б.;

2) термин сөздің немесе терминдік тіркестің басқа тілдегі (орыс тіліндегі) нұсқасы: мысалы, *предложение*, *разделение труда*, *абдукция*, *дисконт* және т.б.;

3) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркесті құрайтын сөздердің этимологиясы және оның қазақ тіліндегі баламасы: мысалы, *сөзжасам* – көне түркі. *сөз*; *жаса* – қазақ; *жаса* + *м* (зат есім тудырушы жұрнақ); *еңбек* – қазақ; *бөл* – қазақ; *бөл* + *ін* (етістік тудырушы жұрнақ) + *іс* (зат есім тудырушы жұрнақ); *абдукция* – лат. *abdūcere*; *ab* – сыртқа қарай, *dūcere* – серпу; *дисконт* – ағыл. *discount* – шегерім, кеміту, жеңілдік және т.б.;

4) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің анықтамасы: мысалы, *сөзжасам* терминінің анықтамасы – *Тіл білімінің жаңа мағыналы туынды сөздердің*

қалыптасуы мен мағыналық дамуын, жасалу тәсілдерін зерттейтін саласы; еңбек бөлінісі терминінің анықтамасы – *Еңбек қарекеті түрлерінің аражігін ажырату; еңбекті ұйымдастыру жүйесінде қалыптасады; өндірістің тиімділігін арттырудың тұрақты қолданыстағы факторы. Е.б.түрлері: жалпы Е.б., жеке Е.б., жеке-дара Е.б., аумақтық Е.б., халықаралық Е.б., атқарымдық Е.б., технологиялық Е.б.; кәсіптік-біліктілік Е.б.; абдукция терминінің анықтамасы – Омыртқалы жануарлар мен адам денелерін омыртқа жотасының бойымен өтіп, оларды тепе-тең екі жарты бөлікке бөлетін орталық сагиттальды жазықтықтан сыртқа (оң немесе сол жаққа) қарай бағытталған қозғалыс; дисконт терминінің анықтамасы – Биржалық және валюталық мәмілелерде тауар, валюта бағасынан шегерім; банк вексельді есептеу кезінде (банктің вексель бойынша төлем төлеу мерзімі басталғанға дейін вексель иелерінен оны сатып алуы кезінде) өндіріп алатын есептік пайыз. Осы сәттегі баға мен өтеу сәтіндегі баға немесе бағалы қағаздың атаулы құны арасындағы, форвардтық баға (белгілі бір мерзімде өтелеетін форвардтық операциялар жүзеге асырылатын мерзімдік мәміле бағасы) мен валютаны дереу жеткізу кезіндегі баға арасындағы, түрлі мерзімде жеткізілетін белгілі бір тауар бағасының арасындағы айырма ретінде айқындалады;*

5) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің анықтамасы алынған дереккөз: мысалы, *сөзжасам* термині анықтамасының дереккөзі – *Н.Оралбай, К.Құрманәлиев, Ж.Балтабаева. Қазақ тілінің сөзжасамы. – Алматы: Інжу-Маржан баспаханасы, 2014. – 10-б.; еңбек бөлінісі* термині анықтамасының дереккөзі – *Қазақ тілі терминдерінің салалық ғылыми түсіндіріме сөздігі. экономика және қаржы. А.: «Мектеп», 2003. – 304 б.; абдукция* термині анықтамасының дереккөзі – *Қазақ тілі терминдерінің салалық ғылыми түсіндіріме сөздігі: Биология. – Алматы: «Мектеп» баспасы», ЖАҚ, 2002. – 339 бет; дисконт* термині анықтамасының дереккөзі – *Қазақ тілі терминдерінің салалық ғылыми түсіндіріме сөздігі. Экономика және қаржы. А.: «Мектеп», 2003. – 304 б.;*

6) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің қолданылатын саласы (салалары): мысалы, *сөзжасам* терминінің қолданылатын саласы – *тіл білімі; еңбек бөлінісі* терминінің қолданылатын саласы – *экономика; абдукция* терминінің қолданылатын саласы – *биология; дисконт* терминінің қолданылатын саласы – *экономика, статистика, стандарттау;*

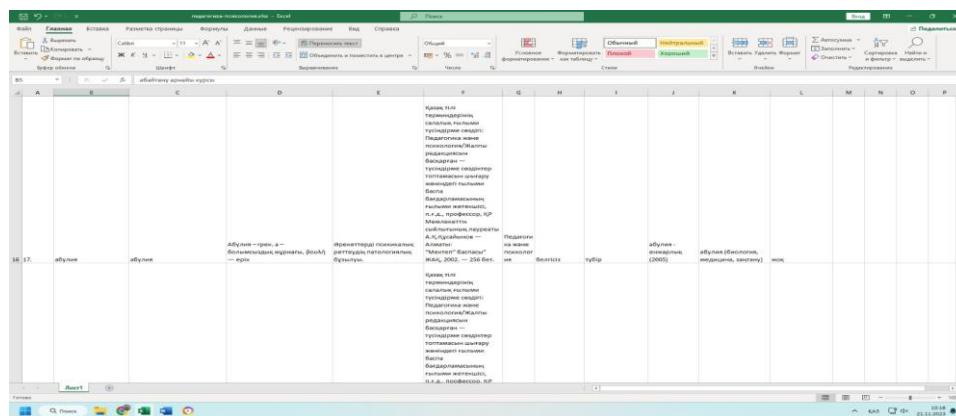
7) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің авторы. Авторы белгісіз терминдерге ешқандай белгі қойылмайды. Ал көптеген төл терминіміздің авторлары белгілі екендігі ақиқат. Мысалы, *септік, рай, шақ* және т.б. терминдердің авторы – *А.Байтұрсынұлы* деп көрсетілді;

8) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің құрамы: мысалы, *абдукция, дисконт* – түбір; *қазына+шы* – туынды; *сөзжасам* – біріккен, *еңбек бөлінісі* – сөз тіркесі және т.б.;

9) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің бекітілген нұсқасы, жылы. Мұнда әрбір термин (терминдік тіркес құрамындағы терминдер) *termincom.kz* сайтынан ізделіп, оның қай жылы қай сөздің баламасы ретінде бекітілгені көрсетілді. Мысалы, *принцип* терминінің бекітілген нұсқалары және жылы: *сөзжасам* (2020); *абдукция* (2018); *жеке еңбек бөлінісі* (2016); *дисконт* (2017); *принцип* (2016); *принцип қағидат* (2018) және т.б.;

10) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің салалық терминологиялық сөздіктердегі аудармасы. Мұнда әрбір термин (терминдік тіркес құрамындағы әрбір термин) 30 томдық салалық терминологиялық сөздіктерден ізделіп, қай саланың сөздігінде қандай нұсқада аударылғаны көрсетіледі. Мысалы, *принцип* терминінің сөздіктердегі аудармасы: *принцип* (экономика, заң); *қағидат* (медицина); *принцип, қағидат* (іс жүргізу) және т.б.;

11) төл және шет тілдерінен енген термин сөздің немесе терминдік тіркестің заңнамадағы қолданысы. Мұнда әрбір термин (терминдік тіркес құрамындағы әрбір термин) *adilet.kz* сайтынан ізделіп, қандай сөзбен аударылғаны көрсетіледі. Мысалы, *сөзжасам* терминінің заңнамадағы қолданысы: *сөзжасам; дисконт* терминінің заңнамадағы қолданысы: *дисконт; еңбек бөлінісі* терминінің заңнамадағы қолданысы: *еңбек бөлінісі; принцип* терминінің заңнамадағы қолданысы: *принцип – принцип, қағидат* және т.б.



2-сурет – ҚТҰК-ның терминологиялық ішкорпусында термин сөздер мен терминдік тіркестерге қойылған терминологиялық белгіленім жүйесінің үлгісі

Корпустық лингвистика саласындағы, оның ішінде терминологиялық ішкорпустарға арналған ғылыми зерттеулер мен қолданбалы құралдарды талдай келе, мынадай тұжырым жасаймыз:

- лексикографиялық корпустар әдетте ауқымды әрі жалпы болса, терминологиялық ішкорпустар нақты пән саласына арналады;
- терминологиялық ішкорпустар терминолог зерттеушілерге тілдік материал жинау барысында уақытты үнемді пайдалануға және осы арқылы зерттеу көлемін, тиімділігін арттыруға мүмкіндік береді;
- терминологиялық ішкорпустар әртүрлі жанрдағы мәтіндерден терминдерді іздеп табу мен қолдануға жәрдемдеседі;
- терминологиялық ішкорпустар терминологиялық тіркесімділік заңдылығын айқындай отырып, синтаксистік және семантикалық анализорлар жасауды, терминологиялық ақпаратты өңдеуді жеңілдетеді;
- терминологиялық ішкорпустардың мазмұны онда қамтылатын ресурстардың қолжетімділігіне, авторлық құқық мәселесіне қатысты шектеулі болуы мүмкін;
- терминологиялық ішкорпустардың көлемі оған енгізілетін мәтіндердің ерекшелігіне (саласына, пәніне, жанрына, типіне және т.б.) байланысты болады;
- терминологиялық ішкорпустардың мүмкіндігі мына жайттарға байланысты болады: терминдерді іздеу жылдамдығы, «ақылдылығы», жабықтығы (статикалығы), яғни аяқталған соң өзгермеуі немесе ашықтығы (динамикалығы), яғни үнемі толықтырылып отыруы және т.б.

ҚТҰК-ның терминологиялық ішкорпусын әзірлеу барысында мынадай нәтижелер алынды:

- терминологиялық ішкорпусына қазақ тіліндегі әртүрлі ғылым салаларына арналған еңбектерден жиналып, сұрыпталып, жүйеленген және техникалық өңдеуден өткізілген, 3 300 000 сөзқолданыстан тұратын ғылыми мәтін енгізілді;
- ҚТҰК-ның терминологиялық ішкорпусының базасына енгізілген ғылыми мәтіндердің әрқайсысына мынадай метамәтіндік белгіленімдер қойылды: Дереккөзі, Мәтін типі, Мәтін тақырыбы, Мәтін стилі, Мәтіннің кілт сөздері, Ғылым саласы, Мәтін тілі, Мәтіннің бет саны, Мәтін графикасы, Сөзқолданыс саны, Аудитория жасы, Субкорпус, Мәтіннің енгізілген күні, Белгілеуші;
- терминологиялық ішкорпусқа енгізілген әртүрлі ғылым салаларының терминдеріне (8000 термин) мынадай параметрлерді қамтитын терминологиялық белгіленім түрлері қойылды: қазақ тіліндегі терминнің орыс тіліндегі нұсқасы, терминнің және терминдік тіркесті құрайтын әрбір бірліктің этимологиясы және шет тілдік терминдер этимологиясының қазақ тіліндегі мағынасы, терминнің анықтамасы, терминнің анықтамасы алынған дереккөз, терминнің қолданылатын саласы (салалары), терминнің авторы, терминнің құрамы, терминнің бекітілген нұсқасы, жылы, терминнің салалық терминологиялық сөздіктердігі аудармасы, терминнің заңнамадағы қолданысы.

Қорытынды

Терминологияның ішкорпус әзірлеудің ғылыми-практикалық ұстанымдарының негізінде жалпы лексикографиялық корпус теориясы жатқанымен, бұл екеуінің мынадай айырмасы болады: лексикографиялық корпустар әдетте кең әрі жалпы болады да, терминологиялық ішкорпустар, керісінше, нақты пән саласына арналады.

Ғылыми (терминологиялық, техникалық немесе арнаулы) дискурсты зерттеуде терминологиялық ішкорпустарды пайдалану лексикалық бірліктердің терминдік мағынасын сәйкестендіруге, олардың көпмағыналылығы мен синонимділігін көрсетуге, олардың қолданыс жілігін анықтауға, олардың арасындағы семантикалық қатынастарды айқындауға мүмкіндік береді.

Терминологиялық ішкорпуспен жұмыс істеу кезінде мына жайттарды ескеру қажет: а) арнаулы ғылым саласы бойынша жазылған мәтіндік материалдың көп мөлшері талап етіледі; ә) әртүрлі ғылым салалары бойынша жазылған қазақ тіліндегі еңбектер саны өте аз; б) терминдер арнаулы мәтіндерден тыс әдетте өте сирек қолданылатын бірліктер болып табылады және белгілі бір ғылым саласына (немесе ұқсас ғылымдар салаларына) арналған мәтіндерде қолданылады; в) қазіргі күн сайын еніп жатқан жаңа терминдер қазақ тіліндегі ғылыми мәтіндерде сирек қолданылуы немесе кездеспеуі мүмкін; г) әдетте терминдердің семантикасы маман еместерге түсініксіз болып келеді.

Терминологиялық ішкорпус мынадай мақсаттарға қызмет етеді:

- терминологиялық ішкорпусты жасау корпус арқылы көпқызметті іздеу жүйесін пайдалануға және терминдердің мәтіндік қормен (мысалдармен) қамтылуына мүмкіндік береді;

- терминологиялық ішкорпустың жасалуы ғылым салалары бойынша зерттеулер жүргізуде, жалпы қолданысты терминдерді бірізді, жүйелі қолдануға ықпал етеді;

- терминологиялық ішкорпус әзірлеу әртүрлі ғылым салалары бойынша аудармаларды сәтті жасауға, машиналық мәтіндік аудармалар мен онлайн сөздіктер жасауға пайдалы болады;

- қоғамдық қатынастарда жиі қолданылатын терминдерді, сонымен қатар ғылым салалары терминдерін (тарсалалық), әсіресе оқулықтар мен оқу құралдарындағы, яғни білім беру жүйесінің терминологиясын біріздендіруге септігін тигізеді;

- ғылым салалары терминдерін жинақтап, тіркеуші, жаңа қолданыстағы терминдерді реттеуші, терминологиялық үдерістерді бақылау қызметін атқарып, сала мамандарының, терминолог-ғалымдардың және т.б., жалпы көпшіліктің терминологиялық білімін жетілдіруге пайдасын тигізіп, мемлекеттік тілдің этносаралық қарым-қатынасты күшейтетін инновациялық, ақпараттық-зияткерлік база ретінде қолданылады.

Зерттеуден алынған нәтижелер теориялық тұрғыда ҚТҰК, оның ішкорпустарын әзірлеудің, әртүрлі терминологиялық сөздіктер (екітілді және көптілді салалық, түсіндірме) құрастырудың ғылыми негіздерін жетілдіруге, қолданбалы терминология, терминография салаларының теориялық базасын дамытуға ықпалын тигізсе, зерттеудің практикалық маңыздылығы әртүрлі ішкорпустар әзірлеу мен терминологиялық сөздіктердің (түсіндірме, реестрлік және т.б.) құрастыру ісінде лексикографиялық практикада қолдануға болатындығымен дәйектеледі.

Мақала ЖТН BR21882249 «Мәдениетаралық коммуникация құралы ретіндегі Қазақ тілінің ұлттық корпусын (ҚТҰК) жетілдіру және оның ішкорпустарын кеңейту» атты бағдарламалық-нысаналы зерттеу аясында жазылды.

Әдебиеттер

Аязбаев Ғ.М., Әшімбаева Н.М., А.З. Бисенғали, Құлманов С.Қ., Нұрлан М. (2023) Қазақ және орыс тілдерінің параллель ішкорпусы: әзірлемесі, қолданылуы және мәселелері // TILTANYM. – № 2 (90) 2023. – бб. 49-61.

Bayekeeva, A., Tazhibayeva, S., Shaheen, A., Beisenova, Z., & Mamayeva, G. (2022). Developing an Online Kazakh-English-Russian Thesaurus of Industry-Specific Terminology. International Journal of Society, Culture & Language, 2022. –№ 10 (1), – P. 83-98.

Бутенко Ю. И., Николаева Н. С., Маргарян Т. Д. (2021) Структурные модели терминологических словосочетаний для разметки корпуса научно-технических текстов // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2021. – Т. 19. – № 3. – С. 45-56.

- Butenko Iu. I., Garazha V. V. (2021). BMSTU Corpus of Scientific and Technical Texts: Conceptual Framework. *Applied Linguistics Research Journal*, 2021. – Vol. 5 (3). – P. 76-81.
- Elena Yu. Balashova, Ekaterina A. Bogacheva, Nikita I. Ilyukhin (2022). Pragmatical, terminological and functional analysis of the un corpus // *Language of Science and Professional Communication*. – 2022. – № 2 (7). – P. 25-43.
- Исаева Г., Отебаева Э., Серікқызы М. (2023) Қазақ тілі ұлттық корпусының лингвомәдени ішкорпусы. *TILTANYM*. – № 3 (91) 2023. – бб. 135-143.
- Jurgita Mikeliūnienė, Jurgita Motiejūnienė (2021). Corpus-based analysis of semi-automatically extracted artificial intelligence-related terminology // *Journal of Language and Cultural Education*. – 2021. – № 9 (1). – P. 30-38.
- Күркімбаева А.М. (2023) «Black Friday» концептісінің жарнама дискурсы прагматикалық сипаты // *TILTANYM*. – № 2 (90) 2023. – 164-173-66.
- Martynova, E. V., Solnyshkina, M. I., Merzlyakova, A. F., Gizatulina, D. Yu. (2020). Lexical Parameters of the Academic Text (Based on the Texts of the Academic Corpus of the Russian Language), *Philology and Culture*, 2020. – № 3. – P. 72-80.
- Сейітбекова А.А., Сейдамаг Ә. (2022) Тарихи ішкорпус әзірлеудің әлемдік тәжірибесінен // *Абылай хан атындағы ҚазХҚ және ӘТУ Хабаршысы. Филология сериясы*. – Том 66. – №3 (2022). – 159-170-66.
- Sergei I. Monakhov, Vladimir V. Turchanenko, Dmitrii N. Cherdakov (2023). Terminology use in school textbooks: corpus analysis // *Research result. Theoretical and Applied Linguistics*, 2023. – № 9 (1). – P. 27-49.
- Tazhibayeva S.Zh., Bayekeyeva A.T. (2023) Digital tools for writing and translation: enhancing language processing and communication in the digital age // *ҒЫЛЫМ. ӘЙЕЛ. ҚОҒАМ*. – 4 (13) 2023. – бб. 166-174.
- Tlegenova Gulden Bakytkazeyevna, Rysbergen Kyzdarkhan Kurmashkyzy, Zhanabekova Aiman Abdildayevna (2024). Representativeness of Mythonyms and Theonyms in Language Corpora // *International Journal of Society, Culture, & Language*. – 12 (1), 2024. – P. 374-395.
- Zhanabekova A.A., Kozhakhmetova A.K., Tlegenova G.B. (2022) Technology for preparing a dialectological internal national corpus of the kazakh language // *Абылай хан атындағы ҚазХҚ және ӘТУ Хабаршысы*. – № 3 (66) 2022. – бб. 70-87.
- Zhanna Sadirmekova, Madina Sambetbayeva, Sandugash Serikbayeva, Gauhar Borankulova, Aigerim Yerimbetova, Aslanbek Murzakhmetov (2023) // Development of an intelligent information resource model based on modern natural language processing methods // *International Journal of Electrical and Computer Engineering (IJECE)* 2023. – Vol. 13. – № 5. – P. 5314-5332.
- Zeinep M. Bazarbayeva, Sholpan K. Zharkynbekova, Aisaule Zh. Amanbayeva, Zhanar T. Zhumabayeva, Ainur A. Karshygyayeva (2023) // The National Corpus of Kazakh Language: Development of Phonetic and Prosodic Markers. *Journal of Siberian Federal University. Humanities & Social Sciences*. 2023. – 16 (8). – P. 1256-1270.

References

- Aiazbaev G.M., Ashimbaeva N.M., Bissengali A.Z., Kylmanov S.K., Nyrlan M. (2023) Qazaq jane orys tilderinin parallel' ishkorpysy: azirlemesi, qoldanylyy jane maseleleri. – *TILTANYM*. – № 2 (90) 2023. – bb. 49-61. [Ayazbayev G.M., Ashimbayeva N.M., Bissengali A.Z., Kulmanov S.K., Nurlan M. (2023) Parallel subcorpus of the Kazakh and Russian languages: development, application and problems. – *TILTANYM*. – № 2 (90) 2023. – P. 49-61. (in Kazakh)]
- Bayekeyeva, A., Tazhibayeva, S., Shaheen, A., Beisenova, Z., & Mamayeva, G. (2022). Developing an Online Kazakh-English-Russian Thesaurus of Industry-Specific Terminology. – *International Journal of Society, Culture & Language*. – 10 (1), P. 83-98. (in English)
- Bytenko Ju. I., Nikolaeva N. S., Margarian T. D. Stryktyrnye modeli terminologicheskikh slovosochetaniy dlya razmetki korpysa naychno-tehnicheskikh tekstov. – *Vestnik NGU. Seriya: Lingvistika i mezhklytynnaia kommunikatsiya*. – 2021. – T. 19, № 3. – S. 45-56. [Butenko Ju. I., Nikolayeva N. S., Margaryan T. D. Structural models of terminological phrases for marking up the corpus of scientific and technical texts. *Bulletin of the NSU. Series: Linguistics and Intercultural Communication*. – 2021. – T. 19, № 3. – P. 45-56. (in Russian)]
- Butenko Iu. I., Garazha V. V. (2021). BMSTU Corpus of Scientific and Technical Texts: Conceptual Framework. *Applied Linguistics Research Journal*. – Vol. 5 (3). – P. 76-81. (in English)
- Yelena Yu. Balashova, Yekaterina A. Bogacheva, Nikita I. Ilyukhin (2022). Pragmatical, terminological and functional analysis of the un corpus. – *Language of Science and Professional Communication*. – № 2 (7). – P. 25-43. (in English)
- Isaeva G., Otebaeva E., Serikqyzy M. (2023) Qazaq tili ulttyq korpysynyn lingvomadeni ishkorpysy. – *TILTANYM*. – № 3 (91) 2023. – bb. 135-143. [Issayeva G., Otebayeva E., Serikqyzy M. (2023) The linguistic and cultural subcorpus of the national corpus of the Kazakh language. – *TILTANYM*. – № 3 (91) 2023. – P. 135-143. (in Kazakh)]
- Jurgita Mikeliūnienė, Jurgita Motiejūnienė (2021). Corpus-based analysis of semi-automatically extracted artificial intelligence-related terminology. – *Journal of Language and Cultural Education*. – № 9(1). – P. 30-38. (in English)
- Kurkimbaeva A.M. (2023) «Black Friday» konceptisinin jarnama diskursy pragmatikalıq sıpaty. – *TILTANYM*. – № 2 (90) 2023. – bb. 164-173. [Kurkimbayeva A.M. (2023) The pragmatic nature of the advertising discourse of the «Black Friday» concept. – *TILTANYM*. – № 2 (90) 2023. – P. 164-173. (in Kazakh)]
- Martynova, E. V., Solnyshkina, M. I., Merzlyakova, A. F. and Gizatulina, D. Yu. (2020). Lexical Parameters of the Academic Text (Based on the Texts of the Academic Corpus of the Russian Language), *Philology and Culture*, 3. – P. 72-80. (in English)

Seitbekova A.A., Seidamat A. (2022) Tarihi ishcorpys azirleydin alemdik tajiribesinen. Abylai han atyndagy QazHQzhaneATU Habarshysy. Filologia seriasy. – T. 66. – № 3 (2022). – bb. 159-170. [Seitbekova A.A., Seidamat A. From the world experience of developing a historical subcorpus. Bulletin of Ablai Khan KazUIRandWL, Philological sciences. – V. 66. – № 3 (2022). – P. 159-170.] (in Kazakh)

Sergei I. Monakhov, Vladimir V. Turchanenko, Dmitrii N. Cherdakov (2023). Terminology use in school textbooks: corpus analysis. – Research result. Theoretical and Applied Linguistics, – № 9 (1). – P. 27-49. (in English)

Tazhibayeva S.Zh., Bayekeyeva A.T. (2023) Digital tools for writing and translation: enhancing language processing and communication in the digital age. – Science. Woman. Society. – 4 (13) 2023. – P. 166-174. (in English)

Tlegenova G.B., Rysbergen K.K., Zhanabekova A.A. (2024) Representativeness of Mythonyms and Theonyms in Language Corpora // International Journal of Society, Culture, & Language, 12 (1), 2024. – P. 374-395. (in English)

Zhanabekova A.A., Kozhakhmetova A.K., Tlegenova G.B. (2022) Technology for preparing a dialectological internal national corpus of the kazakh language. – Bulletin of Ablai Khan KazUIRandWL. – № 3 (66) 2022. – P. 70-87. (in English)

Zhanna Sadirmekova, Madina Sambetbayeva, Sandugash Serikbayeva, Gauhar Borankulova, Aigerim Yerimbetova, Aslanbek Murzakhmetov (2023) Development of an intelligent information resource model based on modern natural language processing methods. – International Journal of Electrical and Computer Engineering (IJECE). – Vol. 13. – № 5. – P. 5314-5332. (in English)

Zeinep M. Bazarbayeva, Sholpan K. Zharkynbekova, Aisaule Zh. Amanbayeva, Zhanar T. Zhumabayeva, Ainur A. Karshygayeva (2023) The National Corpus of Kazakh Language: Development of Phonetic and Prosodic Markers. Journal of Siberian Federal University. Humanities & Social Sciences. 2023. – 16 (8): P. 1256-1270. (in English)