

**Балжан Қанатқызы Қамыза¹, Тыныштық Нұрдаулетқызы Ермекова^{2*},
Гульнар Амангельдиевна Ескермесова³**

¹докторант, Қазақ ұлттық қыздар педагогикалық университеті, Қазақстан, Алматы қ.,
ORCID: 0000-0002-7760-5564 E-mail: kamyzanova.balzhan@gmail.com

^{2*}автор-корреспондент, филология ғылымдарының докторы, профессор,
Қазақ ұлттық қыздар педагогикалық университеті, Қазақстан, Алматы қ.,
ORCID: 0000-0003-1184-2679 E-mail: yermekovat@gmail.com

³философия докторы (PhD), Қазақ ұлттық қыздар педагогикалық университеті,
Қазақстан, Алматы қ., ORCID: 0000-0001-7011-8862 E-mail: Gulnar_9.12.81@mail.ru

СОМАТИКАЛЫҚ КОМПОНЕНТІ БАР ЭТНОБІРЛІКТЕРДІҢ ЖИ ЖҮЙЕЛЕРІНДЕГІ КОГНИТИВТІК МОДЕЛЬДЕНУІ МЕН АУДАРМАДАҒЫ СЕМАНТИКАЛЫҚ АУЫТҚУЛАР

Аңдатпа. Мақала қазіргі компьютерлік лингвистика мен NLP саласындағы өзекті мәселе – ірі тілдік модельдердің (LLM) қазақ тіліндегі этномәдени кодтар мен ұлттық дүниетанымдық қабаттарды тану қабілетін зерттеуге арналған. Өйткені қазіргі қолданыста ЖИ-ге негізделген аударма жүйелеріндегі негізгі олқылықтың басым бөлігі – этномәдени бірліктерді дұрыс тани алмауында болып отыр, сонымен бірге дәл осы салаға қатысты зерттеулердің тапшылығы, мәдени мағыналарды дұрыс жеткізе алатын аударма эрратологиясына бейімді ЖИ үлгілерін жасауға деген сұраныстың артуы – зерттеудің өзектілігін көрсетеді. Зерттеудің мақсаты – қазақ соматикалық этнографизмдерінің (жүрек жалғау, бауырымдау, аузына түкірту, т.б.) ChatGPT, Gemini, Claude жүйелерінде өңделуінің когнитивті механизмдерін айқындау және цифрлық ортадағы семантикалық ауытқуларды (аберрацияларды) жіктеу. Жұмыста когнитивті модельдеу, лингвомәдени талдау және салыстырмалы-эксперименттік әдістер кешені қолданылды. Эмпирикалық база ретінде Ә. Кекілбаев, М. Әуезов, Т. Әлімқұлов және Т. Ахтановтың классикалық шығармаларынан алынған соматикалық этнобірліктер іріктелді. Модельдердің контекске сезімталдығын тексеру үшін екі кезеңдік (лексикалық және контекстік-прагматикалық) лингвистикалық промпттар әзірленді. Эксперименттік талдау нәтижесі ЖИ жүйелерінің қазақ этносоматикалық кодтарын түсіндіруде «прагматикалық инверсия», «лингвомәдени деритуалдау» және «денотаттық буквализм» сияқты қателіктерге жиі бой алдыратынын көрсетті. Модельдерде сакралды мағыналарды батыстық этикалық нормалармен алмастыру және когнитивті инерция құбылысы байқалды. Жұмыстың құндылығы – генеративті қосымшалардың (ChatGPT, Gemini, Claude) этномәдени бейімсіздігін қазақ тілі материалында алғаш рет когнитивті-семантикалық тұрғыдан қарастыруында. Зерттеу нәтижелерінің практикалық маңызы ЖИ модельдерін қазақ тілінің этнографиялық деректер қоры негізінде мамандандырылған түрде баптауда (fine-tuning), сондай-ақ машиналық аударма сапасын арттыруда қолдану мүмкіндігімен айқындалады.

Тірек сөздер: этнографизм; жасанды интеллект (ЖИ); танымдық модельдеу; семантикалық ауытқу; лингвомәдени деритуалдау; аударма.

Қаржыландыру көзі: Мақала Қазақстан Республикасы Ғылым және жоғары білім министрлігі Ғылым комитетінің AP26105015 «Этномәдени бірліктердің жасанды интеллект жүйесінде берілуі: тілдік ауытқулар және оны жүйелеу» атты гранттық қаржыландыру жобасы аясында әзірленді.

Сілтеме жасау үшін: Қамыза Б.Қ., Ермекова Т.Н., Ескермесова Г.А. Соматикалық компоненті бар этнобірліктердің ЖИ жүйелеріндегі когнитивтік модельденуі мен аудармадағы семантикалық ауытқулар. *Tiltany*, 2026. №3 (103). 212-229-66.

DOI: <https://doi.org/10.55491/2411-6076-2026-3-212-229>

**Балжан Канатқызы Камыза¹, Тыныштық Нурдаулетовна Ермекова^{2*},
Гульнар Амангельдиевна Ескермесова³**

¹докторант, Казахский национальный женский педагогический университет,
Казахстан, г. Алматы, ORCID: 0000-0002-7760-5564 E-mail: kamyzanova.balzhan@gmail.com

^{2*}автор-корреспондент, доктор филологических наук, профессор,
Казахский национальный женский педагогический университет, Казахстан, г. Алматы,
ORCID: 0000-0003-1184-2679 E-mail: yermekovat@gmail.com

³доктор философии (PhD), Казахский национальный женский педагогический университет,
Казахстан, г. Алматы, ORCID: 0000-0001-7011-8862 E-mail: Gulnar_9.12.81@mail.ru

КОГНИТИВНОЕ МОДЕЛИРОВАНИЕ ЭТНОЕДИНИЦ С СОМАТИЧЕСКИМ КОМПОНЕНТОМ В СИСТЕМАХ ИИ И СЕМАНТИЧЕСКИЕ ДЕВИАЦИИ В ПЕРЕВОДЕ

Аннотация. Статья посвящена актуальной проблеме современной компьютерной лингвистики и NLP – исследованию способности крупных языковых моделей (LLM) распознавать этнокультурные коды и национально-мировоззренческие пласты казахского языка. Актуальность исследования обусловлена тем, что основным пробелом в существующих системах перевода на базе искусственного интеллекта является неспособность адекватно распознавать этнокультурные единицы. Кроме того, наблюдается дефицит исследований в данной области и растет спрос на создание моделей ИИ, адаптированных к переводческой эрратологии и способных точно передавать культурные смыслы. Цель исследования – выявление когнитивных механизмов обработки казахских соматических этнографизмов (*жүрек жалғау, бауырымдау, аузына түкірту* и др.) в системах ChatGPT, Gemini и Claude, а также классификация семантических отклонений (аббераций) в цифровой среде. В работе использован комплекс методов: когнитивное моделирование, лингвокультурный анализ и сравнительно-экспериментальный метод. В качестве эмпирической базы отобраны соматические этноединицы из классических произведений А. Кекилбаева, М. Ауэзова, Т. Алимкулова и Т. Ахтанова. Для проверки чувствительности моделей к контексту были разработаны двухэтапные (лексические и контекстно-прагматические) лингвистические промпты. Результаты экспериментального анализа показали, что системы ИИ при интерпретации казахских этносоматических кодов часто допускают такие ошибки, как «прагматическая инверсия», «лингвокультурная деритуализация» и «денотативный буквализм». В моделях были выявлены явления подмены сакральных смыслов западными этическими нормами и когнитивная инертность. Ценность работы заключается в том, что в ней впервые на материале казахского языка в когнитивно-семантическом аспекте рассмотрена этнокультурная дезадаптация генеративных приложений (ChatGPT, Gemini, Claude). Практическая значимость результатов исследования определяется возможностью их применения для специализированной настройки моделей ИИ (fine-tuning) на основе этнографической базы данных казахского языка, а также для повышения качества машинного перевода.

Ключевые слова: этнографизм; искусственный интеллект (ИИ); когнитивное моделирование; семантическое отклонение; лингвокультурная деритуализация; перевод.

Источник финансирования: Статья подготовлена в рамках проекта грантового финансирования Комитета науки Министерства науки и высшего образования Республики Казахстан AP26105015 «Трансмиссия этнокультурных единиц в системе искусственного интеллекта: языковые девиации и ее систематизация».

Для цитирования: Камыза Б.К., Ермакова Т.Н., Ескермесова Г.А. Когнитивное моделирование этноединиц с соматическим компонентом в системах ИИ и семантические девиации в переводе. *TiltanyM*, 2026. №3 (103). С. 212-229. (на каз. яз.)

DOI: <https://doi.org/10.55491/2411-6076-2026-3-212-229>

Balzhan Kamyza¹, Tynyshtyk Yermekova^{2*}, Gulnar Yeskermessova³

¹Doctoral Student, Kazakh National Women's Teacher Training University, Kazakhstan, Almaty,
ORCID: 0000-0002-7760-5564 E-mail: kamyzanova.balzhan@gmail.com

^{2*}Corresponding Author, Doctor of Philology, Professor,
Kazakh National Women's Teacher Training University, Kazakhstan, Almaty,
ORCID: 0000-0003-1184-2679 E-mail: yermekovat@gmail.com

³Doctor of Philosophy (PhD), Kazakh National Women's Teacher Training University,
Kazakhstan, Almaty, ORCID: 0000-0001-7011-8862 E-mail: Gulnar_9.12.81@mail.ru

COGNITIVE MODELING OF ETHNOCULTURAL UNITS WITH SOMATIC COMPONENTS IN AI SYSTEMS AND SEMANTIC DEVIATIONS IN TRANSLATION

Abstract. The article is devoted to a pressing issue in modern computational linguistics and NLP – investigating the ability of Large Language Models (LLMs) to recognize ethnocultural codes and national worldview layers in the Kazakh language. The relevance of this study is driven by the fact that the primary limitation in current AI-based translation systems lies in their inability to correctly identify ethnocultural units, combined with a scarcity of research in this field and a growing demand for AI models adapted to translation erratology capable of accurately conveying cultural meanings. The purpose of the study is to identify the cognitive mechanisms underlying the processing of Kazakh somatic ethnographisms (*zhürek zhalgau, bauyrymdau, auzy na tükirtu*, etc.) across ChatGPT, Gemini, and Claude systems, as well as to classify semantic deviations (aberrations) in the digital environment. The study employs a set of methods, including cognitive modeling, linguocultural analysis, and comparative-experimental techniques. The empirical base comprises somatic ethnounits extracted from the classical literary works of A. Kekilbayev, M. Auezov, T. Alimkulov, and T. Akhtanov. To test the models' context sensitivity, two-stage (lexical and context-pragmatic) linguistic prompts were developed. The experimental analysis results demonstrated that AI systems are frequently prone to errors such as “pragmatic inversion,” “linguocultural deritualization,” and “denotative literalism” when interpreting Kazakh ethnosomatic codes. The models exhibited

phenomena of replacing sacred meanings with Western ethical norms and cognitive inertia. The value of the study lies in its being the first to examine the ethnocultural maladaptation of generative applications (ChatGPT, Gemini, Claude) from a cognitive-semantic perspective using Kazakh language material. The research findings possess practical significance for the specialized fine-tuning of AI models using the Kazakh language's ethnographic database and for improving machine translation quality.

Keywords: ethnographism; artificial intelligence (AI); cognitive modeling; semantic deviation; linguocultural deritualization; translation.

Funding: 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 AP26105015 “Representation of ethnocultural units in the artificial intelligence system: linguistic aberrations and their systematisation”.

For citation: Kamyza, B., Yermekova, T., Yeskermessova, G. Cognitive Modeling of Ethnocultural Units with Somatic Components in AI Systems and Semantic Deviations in Translation. *Tiltanyim*, 2026. No.3 (103). Pp. 212-229. (in Kazakh)

DOI: <https://doi.org/10.55491/2411-6076-2026-3-212-229>

Кіріспе

Қазіргі ірі тілдік модельдердің (LLM) табиғи тілді өңдеудегі (NLP) жоғары көрсеткіштеріне қарамастан, олардың мәтінді шынайы түсіну қабілеті әлі де ғылыми дискурстың өзегінде. Зерттеуші *E. Bender мен A. Koller* зерттеулеріне сүйенсек, тек тілдік формалардың статистикалық жиілігіне негізделген модельдердің таңба мен шынайы әлемдегі мағына арасындағы когнитивті байланысты өздігінен игеруі теориялық тұрғыдан мүмкін емес (Bender, Koller, 2020). Бұл іргелі шектеу ЖИ-дің қазақ тілі сияқты этномәдени коды бай тілдердегі терең концептуалды қабаттарды өңдеу қабілетіне тікелей әсер етеді. Десе де Әл-Фараби атындағы Қазақ ұлттық университеті мен Ахмет Байтұрсынұлы атындағы Тіл білімі институты әзірлеп жатқан үлкен тілдік модельдің (YTM) жоғары сапалы және неғұрлым дәл шешімдерге қол жеткізу мүмкіндігінің болуы бұл модельдің қазақ мәдениетін шынымен «түсіне» алатынына сенімділік береді (Жанабекова, Усенов, Тлегенова, 2026).

Заманауи тіл білімінде ЖИ жүйелерінің ұлттық дүниетанымды бейнелейтін когнитивті құрылымдарды, соның ішінде соматикалық этнобірліктерді өңдеу қабілетін зерттеу – өзекті мәселе. Бұл жөнінде зерттеуші *A. Ислам* мақаласында да нақтыланған: медиа саласындағы аудармалардың сапасы мен мәдени-семантикалық нюанстардың сақталу деңгейі әлі де терең зерттеуді қажет етеді (Ислам, 2025). Бұл мәселе әсіресе соматикалық фразеологизмдер мен этнобірліктерді аударуға келгенде анық байқалады, өйткені мұндай тіркестердің мәні жай ғана сөздердің бірігуі емес, ұлттық ойлау мен қалыптасқан дүниетаным нәтижесінде пайда болады.

ЖИ үлгілеріндегі мағыналық бұрмаланудың тағы бір себебін Yong Cao және т.б. ғалымдар көрсеткен «мәдени сәйкессіздік» негіздемесімен түсіндіруге болады, яғни аталған зерттеушілер ЖИ-ді оқыту базасындағы ағылшын тілінде берілген деректердің басым болуы мәдени ерекшеліктерді қалыс қалдырып, ЖИ өз жауаптарын америкалық стандартқа бейімдейтінін дәлелдеп көрсетті (Cao et al., 2023). Мұның әсерінен қазақ тіліндегі де төл тіркестер мен ғұрыптық нышандар ЖИ тарапынан батыстық прагматикалық үлгіге күштеп телінетіні бар. Сондай-ақ, модельдердің мәдени мәтіндерді қабылдауда этикалық әрі қауіпсіздік сүзгілері де кедергі келтіреді. Осы тұста М. Текгюрлердің зерттеуіне илансақ, Gemini моделі тарихи мәтіндерді контекстен бөліп, оларды қате блокқа қоюға бейім екенін растайды (Tekgürler, 2025). Осы сияқты семантикалық өңдеудегі техникалық кедергі зерттеу нысанымызға алынған «аузына түкіру», «бауырымдау» сияқты экспрессивті-этнографиялық бірліктерді өңдеу кезінде де байқалды. ЖИ аталған бірліктерді ритуалдық құндылық емес, физикалық не агрессивті әрекет ретінде қате танып тұр. Мұндай қайшылықтарды сараптау қазақ тілінің цифрлық кеңістіктегі мағыналық тұтастығын сақтау үшін аса маңызды болмақ.

Мақаланың зерттеу нысаны ретінде қазақ тіліндегі соматикалық құрамдас лингвомәдени бірліктердің жасанды интеллект модельдеріндегі (ChatGPT, Gemini, Claude) көрінісі алынды. Ал зерттеу мақаласының пәні – соматикалық этнобірліктердің ЖИ жүйесіндегі танымдық нормалары мен аударма кезінде туындайтын мағыналық ауытқулар.

Мақаланың зерттеу мақсаты: ЖИ үлгілерінің қазақ тіліндегі соматикалық құрамды этноұғымдарды өңдеу мүмкіндігін тілдің танымдық негізінде талдай отырып мәдени-семантикалық ауытқу себептерін теориялық пайымдау. Зерттеуіміздің міндеттері ретінде

мыналарды көрсетеміз: ЖИ модельдеріндегі (ChatGPT, Gemini, Claude) дене мүшелеріне қатысты этноұғымдардың танымдық құрылымдарын анықтау; қазақ тіліндегі тұрақты атаулар мен ғұрыптық символдарды кодтау кезіндегі «мәдени сәйкессіздікті» пайымдау; мағыналық бұрмалану мен техникалық кедергі белгілерін эмпирикалық тәжірибе негізінде растау. Зерттеу кезінде танымдық модельдеу, салыстырмалы-типологиялық талдау жүргізу, лингвомәдени түсінік беру әдісі, сонымен қоса ЖИ үлгілерімен жүргізілген тәжірибелік тексеру қолданылды.

Зерттеуіміздің болжамы былай: генеративті тілдік модельдердің (ChatGPT, Gemini, Claude) соматикалық фразеологизмдерді дәл интерпретациялауы олардың ұлттық-мәдени контексті және тілдік мағынаны қаншалықты толық меңгергеніне тікелей тәуелді. Ал модельдердегі семантикалық ауытқулар тілдік базаның мардымсыздығынан туындайды. Сондықтан, база қазақ тіліндегі мәдени контексті қаншалықты терең меңгерсе, соматикалық фразеологизмдерді соғұрлым дәл әрі қатесіз түсіндіреді.

Материал және әдістер

Зерттеудің ғылыми-теориялық базасы мен әдіснамалық негізін қазақ тіл біліміндегі когнитивтік лингвистика, этнолингвистика және лингвоаксиология бағытының іргелі тұжырымдары (Қайдар, 2013; Сыздық, 2025; Манкеева, 2014; Смағұлова, 2020 т.б.) құрайды. Сондай-ақ, заманауи генеративті жүйелердегі ұлттық коннотация мен этнореалияларды өңдеу мәселелері бойынша отандық зерттеулердің әдіснамалық ұстанымдары басшылыққа алынды (Жанабекова, Усенов, Тлегенова, 2026).

Тәжірибелік-экспериментальды жұмыс барысында жинақталған соматикалық бірліктердің кең ауқымы (500-ден астам фразеологизм) ChatGPT (GPT-4), Gemini және Claude сияқты заманауи генеративті жүйелердің сынақтарынан тұтастай өткізілді. Алынған нәтижелер көрсеткендей, қазіргі ЖИ модельдерінің тілдік мүмкіндіктері жоғары болғандықтан, бірліктердің басым бөлігін (жиі қолданылатын тіркестерді) аударуда қателіктер жібермеді немесе мағынасын жақын деңгейде бере алды. Алайда, зерттеудің басты мақсаты модельдердің жалпы аударма жасай алу қабілетімен қоса, ұлттық-мәдени кодтар мен терең этнопсихологиялық коннотациялардың тоғысындағы *«соқыр тұстарды»* (лакуналарды) анықтау. Жүргізілген эксперимент барысында модельдердің жалпылама деңгейде қателеспегенімен, дәл осы терең когнитивті-ассоциативті жүктемесі бар, семантикасы физиологиялық емес әлеуметтік-психологиялық астарға негізделген ұғымдарда (оңайлықпен аударылмайтын) қайталана сүрінетіні анықталды.

Дәл осы себептен, барлық стандартты мысалдарды тізіп шықпай, модельдердің осал тұстарын, яғни лингвомәдени лакуналарды нақты айқын көрсететін төрт этноұғым (*«жүрек жалғау»*, *«қу тізесін құшақтау»*, *«ой бауырымдап келу»*, *«аузына түкірту»*) талдауға іріктеп алынды. Бұл тұста мақаланың көлемдік шектеулері мен әрбір бірлікке терең когнитивті-семантикалық және прагматикалық талдау жасау қажеттілігі ескеріліп, тек 4 түйінді соматикалық этноұғым таңдалды. Бұл тәсіл заманауи ЖИ үлгілерінің тіпті ең озық нұсқаларының өзінде ұлттық коннотацияны өңдеуде қандай шектеулерге ұшырайтынын дәлелді түрде көрсетуге мүмкіндік береді.

Зерттеудің эмпирикалық базасын қалыптастырған бұл тілдік бірліктер – қазақ көркем әдебиетінің классикалық туындыларынан (М. Әуезов, Ә. Кекілбаев т.б.) жиналған танымдық жүктемесі мен қолданылу жиілігі жоғары соматикалық этноұғымдар. Көрсетілген соматизмдерді іріктеу барысында олардың функционалдық бөлімдерін, яғни метафоралық бейнесі, эмоциялық қыры, танымдық сипаты, символдық және әрекетке қатысты белгілерін негізге алдық (Alimjanova, 2025: 76). Соматизмдердің лексикалық мағынасын нақтылау үшін «Қазақ әдеби тілінің сөздігі» және «Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі» атты энциклопедиялық қор қолданылды.

Қазіргі таңда қолданушылардың дәстүрлі машиналық жүйелерден (Google, Yandex) генеративті ірі тілдік модельдерге (LLM) жаппай бет бұруы байқалады. Егер дәстүрлі жүйелер ұлттық-мәдени реңкті жоғалтса, генеративті модельдер контексті ескеріп, этнолексемалардың астарлы мағынасын жеткізуге ұмтылады (Ескермесова, Ермакова, Шойбекова, 2026: 67). Сол себепті зерттеу жұмысы барысында тек генеративті модельдер таңдалып алынды: ChatGPT

(OpenAI өнімі, GPT-4 нұсқасы), Gemini (Google өнімі) және Claude (Anthropic өнімі). Осы тұрғыда ChatGPT, Gemini және Claude модельдерінің зерттеу нысаны ретінде таңдалу себебі: олардың қазақ тіліндегі контентті өңдеудегі кең қолданысымен, функционалдық мүмкіндіктерімен және салыстырмалы талдауға лайықты репрезентативтілігімен негізделеді. Бұл модельдердің қателіктерін сараптау ұлттық-мәдени кодтар мен соматикалық бірліктерді аударудағы ЖИ мүмкіндіктері мен шектеулерін жан-жақты бағалауға мүмкіндік береді.

GPT-4 нұсқасы эксперимент нәтижелерінің тұрақтылығы мен қайталануын қамтамасыз ететін тұрақты базалық үлгі ретінде таңдап алынды. Бұл шешім кейінгі кезеңде жаңарып шығатын модельдердің эволюциялық динамикасын салыстырып тексеру үшін болашақ зерттеушілерге объективті ғылыми орын қалдыруға негіз болды.

Зерттеу жұмысының эксперименттік-сараптамалық бөлігі өзара байланысты үш кезеңнен тұрды: *Бірінші кезең (денотативті талдау)*: бұл сатыда ЖИ үлгілерінің (ChatGPT, Gemini, Claude) тілдік бірліктердің мағыналық көлемін бастапқы деңгейде тану қабілетін тексеру мақсатында соматикалық этнобірліктер контекстсіз (оқшау түрде) беріліп, орыс және ағылшын тілдеріне аудару тапсырылды. Бұл кезеңде модельдерге *«Берілген қазақ тіліндегі тұрақты тіркестің/этнобірліктің тура мағынасын түсіндіріп, орыс және ағылшын тілдеріне баламасын беріңіз: [этнобірлік]»* деген тікелей промпт берілді. Бұл тәсіл модельдердің сөздік-семантикалық базасы мен тура мағынаны беру деңгейін анықтауға жол ашты.

Екінші кезең (контекстік-прагматикалық қабат): бұл кезеңде соматикалық тіркестер қазақ әдебиеті классиктерінің шығармаларынан алынған нақты мәтіндік фрагменттер құрамында (контекст арқылы) ұсынылды. Модельдерге қойылған нақты нұсқау: *«Мәтін үзіндісіндегі асты сызылған соматикалық этнобірліктің көркемдік астары мен этномәдени мәнін сақтай отырып, орыс және ағылшын тілдеріне көркем аударма жасаңыз: [мәтін фрагменті]»* болды. Мұндағы мақсат – ЖИ үлгілерінің көркем мәтін контексіне сезімталдығын, этномәдени астарды және прагматикалық мәнді ажырата алу қабілетін сараптау болатын.

Үшінші кезең (Интерпретация және салыстырмалы талдау): соңғы бөлімде алынған нәтижелер лингвомәдени бояу мен ұлттық коннотация аясында жан-жақты қарастырылып, әрбір жүйенің (ChatGPT, Gemini, Claude) жауаптары өзара салыстырылды. Бұл кезең модельдердің қай тұста сүрінетінін және ұлттық лакуналарды қалай өңдейтінін нақты көруге мүмкіндік берді.

Жоғарыда көрсетілген зерттеу жобасының мәселені қарастыру бөлімдері мен сараптау кезеңдерінде төмендегі әдістерді басшылыққа алдық. *Танымдық модельдеу*: ЖИ-дің дене-мүшелермен байланысты метафораларды түсіну домендерін анықтау кезінде; *Лингвомәдени талдау*: этнобірліктердегі сакралды мағыналардың сақталуын тексеру барысында; *Салыстырмалы-салғастырмалы әдіс*: генеративті модельдердің нәтижелерін өзара салыстыру және семантикалық ауытқуларды (аберрацияларды) жіктеу уағында; *Контекстік талдау*: сөздік мағына мен көркем мәтіндегі қолданыстың сәйкестігін бағалау барысында қолданылды. Осы аталған әдістер мен материалдар негізінде жүргізілген эксперименттік жұмыстың нәтижелері ЖИ модельдерінің қазақ тіліндегі этномәдени бірліктерді танудағы елеулі айырмашылықтарын айқындап берді.

Әдебиетке шолу

Қазақ тіл білімінде ұлттық тілді сол тілді тұтынушы халықтың танымымен тікелей сабақтастыра зерттеген Ә. Қайдар (2013), Р. Сыздық (2025), Н. Уәли (2021), Ж. Манкеева (2014), Г. Смағұлова (2020) т.б. ғалымдардың іргелі зерттеулерінен соң соматизмдерді тікелей зерттеу нысанына айналдырған зерттеулер жазылды: Ә. Ташауов (1995) – «Кодекс Куманикустегі» соматикалық атаулардың лексика-семантикалық топтарын, Б. Уызбаева (1994) – соматикалық етістік фразеологизмдерді, М. Абдрахманова (2008) – «Көз» концептісін, А. Нұржанова (2010) – фразеологизмдердің соматикалық кодын т.т. диссертация деңгейінде қорғады. Бұл еңбектердің құндылығы – тарихи және қазіргі тілдік деректер негізінде халқымыздың адамның дене мүшелері арқылы дүниені тану, қабылдау туралы танымын пайымдауға, ұлт болмысын ашуға мүмкіндік беруінде. Соматикалық лексика – ежелгі қабатты құрайтын, семантикалық құрылымы тармақталған және жоғары жиілікті микрожүйе. Соматизмдерді зерттеу нысаны етіп алған қазіргі зерттеулер аталған нысанды заманауи бағытта зерттеуге ден қоюда. А.А. Раманованың

зерттеулерінде соматизмдер лингвомәдениет пен интермәдени коммуникацияның өзекті бөлшегі ретінде қарастырылып, олардың лексикалық-семантикалық өрісі ұлттық дүниетанымды модельдейтін когнитивтік домен ретінде сипатталады (Раманова, 2023). Бұл мәселе, әсіресе, мәдениетаралық коммуникацияда айқын көрініс табады. Аудармалардағы соматизмдерді зерттеушілер қазақ тіліндегі *бет, түс, жүз, ажар, пішін, шырай, өң* сияқты 7 лексемаға орыс тілінде эквиваленті ретінде бір ғана «лицо» лексемасының ұсынылғанын нақты тілдік аударма деректері арқылы дәлелдей келіп: «Ақпарат жеткізу, сигнал беру тұрғысынан адамның беті біріншілік, басқа соматизмдер (қол, аяқ, дене т.б.) көбінесе екіншілік қызмет атқарады», – деген қорытынды жасайды (Иманғазина, 2024).

Бұл бағыттағы этномәдени бірліктерді цифрлық ортада сақтау мен алгоритмдік өңдеу мәселелері ұлттық корпус аясында белсенді талқылануда. А.С. Барменқұлова мен Ж.А. Манкеева этномәдени атауларды кодтау мен модельдеуді ұлттық сана мен мәдени жадты сақтаушы инфрақұрылым ретінде ұсынады (Барменқұлова, Жаңабекова, Манкеева, 2023). Ал Г.С. Исаева мен Э.Ә. Өтебаеваның еңбектерінде лингвомәдени ішкорпус аясында этномаркер бірліктердің мәдени-семантикалық белгіленімін жасау үлгілері берілген (Исаева, Өтебаева, Серікқызы, 2023). Дегенмен бұл теориялық база әлі де болса ірі тілдік модельдердің (LLM) практикалық эксперименттерімен толық интеграцияланбаған.

Қазіргі таңда LLM модельдерінің қазақ тіліндегі мүмкіндіктеріне қатысты сыни көзқарас қалыптасқан. А. Maxutov пен Р. Braslavski жүргізген пилоттық бағалау көрсеткендей, LLM-дер қазақша генеративтік міндеттерде ағылшын тіліне қарағанда айтарлықтай әлсіз (Maxutov, Myrzakhmet, Braslavski, 2024). GPT-4 сияқты озық модельдердің өзінде «англоцентрлік» басымдықтың салдарынан семантикалық жылжу мен мәдени коннотациялардың жоғалуы байқалады.

Бұл мәселенің бар екенін KazMMLU тестілеу қорытындылары да растады: қазіргі жасанды интеллект модельдері қазақша мәтіннің мағынасын түсінуде және арнайы терминдерді қолдануда көп қате жібереді (Togmanov et al., 2025). Белгілі ғалым Г. Маркустың айтуынша, ЖИ модельдерінде «терең түсіну» қабілеті жоқ. Олар тек сөздердің тіркесу жиілігін (статистиканы) ғана есептейді, бірақ қазақ тіліндегі ұлттық мәдениетпен астарласқан сөздерді (мысалы, соматизмдерді) адам сияқты сезініп, саналы түрде (когнитивтік деңгейде) қабылдай алмайды (Marcus, 2020).

Қазіргі кезде тек қазақ тіліне арналған шағын модельдерді (Small-Scale LLMs) жасау қолға алынуда. Бұл бағдарламаларда қазақ тілінің жалғамалы (агглютинативті) болмысы және этностық индикатор, мәдени белгілер ескерілген (Kadyrbek et al., 2025). Алайда, бұл жаңа жүйелердің өзі тәржімалау не мәтін аудару кезінде сөз мәнін өзгертіп, прагматикалық жаңылысуға ұшырайды (Toleu, Tolegen, 2024). Тіпті қатені жөндеуге бағытталған KazPE жүйелері грамматикалық қателерді реттей алғанымен, сөз мазмұнындағы астарлы мәдени мәнді нақты ашып бере алмайды (Karyukin et al., 2024).

Қысқасы, соматикалық этноұғымдардың ЖИ жүйесінде «техникалық соқырлыққа» тап болуы және семантикалық нормадан қалыс қалуын жоспарлы түрде мониторинг жасап таразылайтын метрикалардың жоқтығы – осы зерттеу жұмысының өзекті тұсын көрсетеді.

Нәтижелер және талқылау

Зерттеу объектісі негізінде таңдалған «жүрек жалғау» тіркесінің семантикалық декодталуы екі сатыда зерделенді. Бұл ұғым қазақ тілінің тұрақты тіркестер базасында «тамақтану, аздап тамақ ішу» деген мәнде қолданылады (Қазақ әдеби тілінің сөздігі, 2011а). Сөздікте көрсетілген сипаттамаға сай тіркестің образды мәні биологиялық қарын ашу процесін бәсеңдету актісіне құрылған. Алғашқы кезеңде ЖИ модельдеріннің (ChatGPT, Gemini, Claude) берілген этнобірліктің халықтық болмысын сөз тұлға көрінісінде (контекстеіз) анықтау қабілеті қаралса, келесі бөлімде Ә. Кекілбаевтың «Тасбақаның шөбі» шығармасынан «*Қайынаға, не ішесіз? Жүрек жалғар бірдеңе болса...*» сөйлемі негізінде модельдердің дискурстық деңгейдегі прагматикалық сезіну қабілеті зерделенді. ЖИ жүйелері ұсынған интерпретациялар, аударма баламалары және анықталған семантикалық ауытқулар Кесте 1-де кешенді түрде жүйеленген.

Кесте 1 – «Жүрек жалғау» соматизмінің ЖИ модельдерінде өңделу нәтижелері

Таблица 1 – Результаты обработки соматизма «жүрек жалғау» в моделях ИИ

Table 1 – Results of Processing the Somatic Unit “zhürek zhalgau” in AI Models

ЖИ моделі	Берілген тіркеске жүйе ұсынған түсіндірме	Ағылшын тіліне аударма (English)	Орыс тіліне аударма (Russian)	Нәтиженің сипаттамасы
Chat GPT	Эмоционалды-элеуметтік фразеологизм. Мағынасы: біреуге жылы, шын ниетпен көңіл бөлу, қамқорлық көрсету немесе дос болу сезімін білдіру.	- To open one's heart to someone; - To warm someone's heart; - To show heartfelt care.	- Раскрыть сердце кому-либо; - Показать душевную теплоту; - Проявить искреннюю заботу.	<i>Концептілік алмастыру:</i> Идиоманың этнографиялық мағынасын жаһандық метафоралық стереотиппен (жүрек = сезім) алмастыру.
		- Father-in-law, what will you drink? If there is something to warm the heart...	- Свекор, что будете пить? Если есть что-нибудь для сердца...	<i>Когнитивті инерция:</i> Контекстегі «ішу/жеу» әрекеті мен сыйластық ситуациясына қарамастан, бастапқы қате модельді (жүрек = сезім) сақтау.
Gemini	Тура мағынасы – қарның қатты ашқанда өзегің талмас үшін кішкене ғана тамақ ішіп, аштықты басу. Бұл – уақытша қуат жинау.	- To grab a quick bite; - To have a snack; - To take the edge off hunger; - To tide someone over.	- Заморить червячка; - Перекусить; - Утолить голод (немного); - Подкрепиться.	<i>Сәйкестік:</i> Этнобірлік мағынасының және «жүрек – энергия көзі» танымдық моделінің сақталуы.
		Brother-in-law, what would you like to drink? If there is something to have a quick bite...	Кайнага (брат мужа), что будете пить? Если найдется что-нибудь перекусить (червячка заморить) ...	<i>Сәйкестік:</i> Контекске сай лексикалық баламаны таңдау.
Claude	Мағыналық түсіндірме берген жоқ. Дәлірек аударма үшін медициналық немесе техникалық контекст қажет деп сұрады.	- Heart connection; - Cardiac connection; - Heart junction; - Cardiac junction.	- Сердечное соединение; - Кардиальное соединение; - Соединение сердца.	<i>Денотаттық қате</i> (деметафоралау): Метафораны тікелей анатомиялық/техникалық термин ретінде қабылдап тура аудару.
		Mother-in-law, what would you like to drink? If there's something to comfort the heart...	Тёща, что будете пить? Если есть что-нибудь, что утешит сердце...	<i>Когнитивті бұрмалау:</i> Мағына «анатомиядан» «эмоцияға» ауысса да этнобірлікті таппау. Мағынаны бұрмалап қана қоймай, туыстық атауды (Қайынаға – Тёща) қате аударды.

«Бауырымдау / ой, бауырымдау» этнобірлігін зерттеу барысында ЖИ жүйелерінің қазақ мәдениетіндегі қайғы мен аза тұту ритуалдарын тану деңгейі зерделенді. Қазақ дәстүрінде бұл бірлік өлік жөнелту салтындағы қайғыру жосынын білдіреді: көңіл айтуға келген ағайын-туыс үй маңына жеткенде «ой, бауырымдап!» дауыстай жылап көңіл айтып келеді (Әлімбаев, 2011). Лексикалық деңгейде модельдердің «бауыр» соматизмін тек туыстық қатынас ретінде қабылдау мүмкіндігі тексерілсе, дискурстық деңгейде М. Әуезовтің «Қаралы сұлу» шығармасынан алынған «Бұл шу ауылға қарай ат қойып, «ой бауырымдап» келе жатқан жылқышылар мен қуғыншылар шуы» сөйлемі ұсынылды. Мұндағы мақсат – ЖИ-дің прагматикалық контекст

әсерінен сөз мағынасын «туыстық жақындықтан» «ритуалдық жоқтауға» ауыстыру қабілетін анықтау болды. Талдау барысында модельдердің бұл салтты агрессиялық «айқай» немесе жай ғана «туысты шақыру» ретінде қате интерпретациялауы сынды семантикалық деформациялар анықталды. Нәтижелер Кесте 2-де топтастырылған.

Кесте 2 – «Ой, бауырымдау» этнобірлігінің ЖИ модельдерінде өңделу нәтижелері

Таблица 2 – Результаты обработки этноединицы «ой, бауырымдау» в моделях ИИ

Table 2 – Results of Processing the Ethno-unit “oy, bauyrymdau” in AI Models

ЖИ	Берілген тіркеске жүйе ұсынған түсіндірме	Ағылшын тіліне аударма (English)	Орыс тіліне аударма (Russian)	Нәтиженің сипаттамасы
Chat GPT	Эмоциялық-экспрессивті тіркес, мағынасы: біреуді жақын тарту, бауырындай көру, жылы ілтипат білдіру, кейде аяушылық, жанашырлық реңкі бар.	- to feel brotherly affection; - to treat someone like a brother; - to feel close to someone; - to develop a sense of kinship	- отнестись по-братски; - почувствовать братскую близость; - принять как родного; - проникнуться тёплым, родственным чувством	Мәдени соқырлық: Ритуалдық мағынаның (аза тұту контекстінің) жоғалуы және оның позитивті әлеуметтік байланыс ретінде қате интерпреттенуі.
		This commotion is the sound of the horsemen and pursuers heading toward the village, <i>showing brotherly care and camaraderie.</i>	Этот шум – это крики всадников и преследователей, направляющихся к деревне, <i>проявляющих братскую теплоту.</i>	Когнитивті инерция: Қайғылы/ритуалдық контекске карамастан, позитивті әлеуметтік байланыс моделін сақтап қалу.
Gemini	Бұл ұғым қазақ өркениеті мен руханиятындағы астарлы эмоцияны, татулықты немесе мейірімділікті білдіреді. Бұл тіркестің мәні контекске (куану, ренжу, тандану не демеу болу) байланысты өзгеруі мүмкін.	- Oh, my dear brother! - Oh, brother! - Oh, my man! - Oh, my dear friend! - Oh, boy!	- Ой, братец мой! - Ой, брат ты мой! - Ох, голубчик! - Братишка!	Лингвомәдени деритуалдау: Ритуалдық форманы күнделікті экспрессивті фразамен шатастыру.
		This noise comes from the horsemen and pursuers galloping toward the village, <i>wailing “oh, my brother” in mourning.</i>	Этот шум – крики табунщиков и преследователей, которые скачут во весь опор к аулу с плачем «ой, бауырым» (<i>о, брат мой</i>).	Контекстік репарация: Дискурстық маркерлер көмегімен (ат қою, шабу) этнографизмнің ритуалдық табиғатын қайта тану және семантикалық ауытқуды жоюға тырысу.
Claude	«Бауырымдау» сөзі қазақ тілінде «жақын болу, жылы қарым-қатынас орнату, жанашырлық таныту» мағынасын білдіреді. Сондықтан контекске байланысты нұсқалар да қолданыла алады.	- Compassionate understanding; - Empathetic connection; - Heartfelt sympathy	- Сочувствие; - Сопереживание; - Душевное расположение	Концептілік трансформация: Этнографиялық сакралды мәнді заманауи психологиялық «bonding» терминімен алмастыру.
		This noise was the noise of herdsmen and drovers galloping on horses toward the village, <i>shouting oh, brothers!</i>	Этот шум – шум табунщиков и погонщиков, скачущих на лошадях к аулу, <i>выкрикивая ой, братцы!</i>	Стильдік алашақтық: Қайғылы оқиғаға қатысты жоқтауды жалпы нұсқадағы бейтарап айқай мен дабырға айналдыру.

«Аузына түкірту» этнобірлігін зерттеу барысында ЖИ жүйелерінің қазақ

дүниетанымындағы «қасиет дарыту» және «бата алу» институтының символикалық сипатын тану мүмкіндігі зерделенді. Қазақы ортада бұл бірлік байырғы контагиозды магиялық жосыннан туындаған рәсім: оның мәні ел сыйлаған, атағы шыққан тұлғалардың жақсы қасиеттері жас балаға жұғысты болсын деген игі ниетке негізделген (Әлімбай, 2011). Эксперименттің бірінші кезеңінде модельдердің «түкіру» әрекетінің физиологиялық мағынасы мен мәдени-магиялық астарының аражігін ажырату қабілеті тексерілді. Дискурстық деңгейде Т. Әлімқұловтың «Көкек айы» шығармасынан алынған «*Мен жүрген жерге қазақтар баласын аунатып алады, аузына түкіртеді. Ал, сен тумаңның есімін өзің қоясың...*» үзіндісі модельдерге контекстік бағдар ретінде ұсынылды.

Мұндағы негізгі мақсат – ЖИ жүйелерінің прагматикалық контекст әсерінен нысанның мағынасын «физиологиялық әктіден» «ритуалдық ізгілікке» трансформациялау қабілетін анықтау болды. Талдау барысында модельдердің этнографиялық мазмұнды жалпылама «қамқорлық» ұғымымен алмастыруы (*этнографиялық деперсонализация*) және кей жағдайда мәдени кодты «қорлау» ретінде қабылдауы (*прагматикалық инверсия*) байқалды. Зерттеу нәтижелері мен анықталған семантикалық ауытқулар Кесте 3-те жүйеленген.

Кесте 3 – «Аузына түкірту» этнобірлігінің ЖИ жүйелерінде өңделген нәтижесі

Таблица 3 – Результаты обработки этноединицы «аузына түкірту» в системах ИИ

Table 3 – Results of Processing the Ethno-unit “auzyna tüpirtu” in AI Systems

ЖИ	Берілген тіркеске жүйе ұсынған түсіндірме	Ағылшын тіліне аударма (English)	Орыс тіліне аударма (Russian)	Нәтиженің сипаттамасы
Chat GPT	Қазақ тіліндегі өте күшті этномәдени, символдық тіркес. Мағынасы: біреуді өзіне толық қаратты, ықпалын жүргізді, сөзін екі етпейтін жағдайға жеткізді, кейде абсолютті бедел, сөзге тоқтау мәнінде қолданылады. (Тікелей физиологиялық әрекет емес!) Қазақ дүниетанымында ауыз – сөздің, анттың, биліктің орталығы. Ал түкіру – көне дәстүрде таңбалау, бекіту, ықпал ету символы. Сондықтан «аузына түкірту» – әлеуметтік иерархияда біреудің сөзін заңдай мойындату дегенді білдіреді.	- to have absolute authority over someone; - to make someone obey unquestioningly; - to exert total influence; - to command unquestioned respect	- иметь безоговорочный авторитет; - держать в полном подчинении; - заставить беспрекословно слушаться; - обладать абсолютным влиянием	Мәдени-прагматикалық деформация. Контагиозды магиялық жосынды (жақсылық жұғысты болу) әлеуметтік үстемдік пен бағындыру моделіне алмастыру.
		Wherever I go, Kazakhs cradle the child and place it in their care. But you will name your own child yourself...	Куда бы я ни пошёл, казахи берут ребёнка на руки и приносят под свою опеку. А ты имя своему ребёнку выбираешь сам...	Этнографиялық деперсоналдау. Нақты мәдени-магиялық ритуалды жойып, оны бейтарап жалпы әлеуметтік қамқорлық ұғымына ауыстыру.
Gemini	Аузына түкірту – бұл қазақ мәдениетіндегі терең мағыналы салт. Оның тікелей аудармасы мен мағыналық баламалары контекске байланысты әртүрлі болады.	Мағыналық аудармасы: A traditional ritual of passing on talent/wisdom. Тікелей аудармасы: To symbolically spit into someone's mouth. Түсіндірмелі аударма (Description): A Kazakh custom where a	Мағыналық аудармасы: Благословение через «передачу дара» (традиция, когда уважаемый человек символически плюет в рот ребенку, чтобы	Лингвомәдени бейімдеу. Этнографиялық мағынаның толық сақталуы және мәдени кодтың нысана тілге сәтті бейімделуі, метафоралық

		respected person symbolically spits into a child's mouth to pass on their talent or charisma. <i>Идиомалық баламасы</i> (жуық мағынада): "To pass the torch" (білімін, талантын келесі ұрпаққа беру). "To bestow a blessing" (бата беру/дарыту).	передать свою мудрость или талант). <i>Тікелей аудармасы:</i> Поплевать в рот. Контекст бойынша: «Получить благословение от великого человека» (ұлы адамнан бата алу). «Перенять талант/способности через традицию» (дәстүр арқылы талантты дарыту).	мазмұнмен сәтті алмастыру.
		Wherever I go, Kazakhs roll their children where I stepped and ask me to bless them (spit in the infant's mouth for good luck). But you shall name your own kin yourself...	Там, где я бываю, казахи валяют своих детей (в пыли моих следов), просят благословить (плюнуть в рот младенцу). А ты сам дашь имя своему родственнику...	<i>Экспликация</i> сәйкестік. Мәдени шок тудыруы мүмкін элементті (түкіру) мағыналық түсіндірме арқылы бейімдеп, ритуалдық мәнін сақтау.
Claude	Бұл тіркес қазақ тілінде екі түрлі мағынада қолданылуы мүмкін: тура мағынада (медициналық немесе басқа арнайы жағдайларда); ауыспалы мағынада (қорлау, кемсіту мағынасында).	<i>Тура мағынада:</i> To spit into someone's mouth. <i>Ауыспалы мағынада:</i> To humiliate, To insult, To show contempt	<i>Тура мағынада:</i> Плевать/плюнуть в рот. <i>Ауыспалы мағынада:</i> Унижать, Оскорблять, Проявлять презрение	<i>Семантикалық контаминация</i> (аксиологиялық инверсия). Позитивті ритуалды «түкіру – қорлау» стереотипіне сүйеніп, негативті мағынамен алмастыру.
		Everywhere I go, Kazakhs are oppressed, spat upon. And you yourself name your own offspring...	Везде, где я бываю, казахов притесняют, плюют им в лицо. А ты сам даёшь имя своему потомству...	<i>Прагматикалық инверсия</i> (абберрация). Позитивті мәдени ритуалды әлеуметтік-саяси қорлау және кемсіту актісі ретінде бұрмалап қабылдау.

«Қу тізесін құшақтады» тұрақты тіркесін ЖИ модельдерінде қарастыру кезінде олардың образды ой орамын түсіндіре алу, сонымен қатар халықтың әлеуметтік-топтық құрамына қатысты ишаралауды тану сапасы зерделенді. Қазақ тілінің фразеологиялық жүйесінде бұл бірлік «әйелсіз, бала-шағасыз өмір сүру; қосағынан айырылып үйленбеді, жалғыз қалу» деген мағыналық жүктемеге ие (Қазақ әдеби тілінің сөздігі, 2011б). Эксперимент барысында Т.Ахтановтың «Шырағың сөнбесін» шығармасынан алынған «*Еркек адам әйелім өлді деп қу тізесін құшақтап отыра ма, үй болу керек*» сөйлемі арқылы модельдердің контекстік ақпаратты өңдеу қабілеті тексерілді.

Мұндағы мақсат – ЖИ-дің «қу тізе» тіркесінің көпмағыналылығын (жалғыздық,

ұрпақсыздық) ажырату және нысана тілдегі стильдік сәйкестікті сақтау мүмкіндігін анықтауға бағытталды. Талдау барысында модельдердің фразеологизмді тікелей физикалық әрекет ретінде қабылдауы (*денотаттық буквализм*) және мағынаны материалдық жетіспеушілікке ауыстыруы (*концептуалды алмасу*) сияқты семантикалық ауытқулар тіркелді. Алынған нәтижелер Кесте 4-те жинақталған.

Кесте 4 – «Қу тізесін құшақтады» соматизмінің ЖИ модельдерінде өңделу нәтижелері

Таблица 4 – Результаты обработки соматизма «ку тізесін құшақтады» в моделях ИИ

Table 4 – Results of Processing the Somatic Unit “qu tizesin qushaqtady” in AI Models

ЖИ	Берілген тіркеске жүйе ұсынған түсіндірме	Ағылшын тіліне аударма (English)	Орыс тіліне аударма (Russian)	Нәтиженің сипаттамасы
Chat GPT	Қазақ тіліндегі фразеологизм, мағынасы: барынан айырылды, тақыр кедейленді, дәрменсіз күйге түсті, көбіне <i>үмітсіздік</i> , <i>сорлылық</i> реңкімен колданылады. Қазақ дүниетанымында: тізе – тірек, қозғалыс, әрекет символы тізесін құшақтау – әрекетсіздік, тығырыққа тірелу. Ал қу эпитеті: бос, құр, иесіз деген бағалауыш мән үстейді. Сондықтан бұл тіркес әлеуметтік күйреу мен үмітсіздіктің жинақы бейнесі.	- to be left destitute; - to be reduced to poverty; - to end up penniless; - to be left with nothing;	- остаться ни с чем; - оказаться в полной нищете; - обнищать; - сидеть без средств к существованию	<i>Семантикалық ауысу</i> (концептуалды алмасу). «Ұрпақсыз қалу/жалғыздық» концептісін «Материалдық кедейлік» концептісімен алмастыру.
		A man should not sit lamenting, <i>holding his empty knees</i> because his wife died; he must build a home.	Мужчина не должен сидеть, <i>обнимая пустые колени</i> , потому что жена умерла; он должен создать дом.	<i>Сөзбе-сөз аудару</i> (метафоралық деформалау). «Қу» метафорасын тікелей физикалық сипаттама (empty/пустой) ретінде қабылдап, тілдік нормаға жат тіркес жасау.
Gemini	Бұл қазақ тіліндегі мұңды, бейнелі фразеологизм. Ол көбінесе жалғыз қалу, перзентсіз өту немесе ешкімі жоқ шарасыз күйге түсу дегенді білдіреді. Орыс тілінде бұл тіркестің мағынасын беретін бірнеше нұсқа бар. Ағылшын тілінде бұл	- To be left all alone in the world; - To be left with nothing but one's own shadow; - To be left high and dry. <i>Тікелей сипаттамасы:</i> - To embrace one's bare knees (бұл фраза ағылшын тілді оқырманға «жалғыздықтан өз-өзін құшақтап отыру» деген бейнені елестетеді).	<i>Мағыналық баламасы:</i> Остаться у разбитого корыта. <i>Тікелей мағынасы:</i> Остаться в полном одиночестве немесе коротать век в одиночестве. <i>Эмоционалды бояуы:</i> Остаться бобылем (ер адамға қатысты: жалғыз, отбасысыз қалу).	<i>Функционалды эквиваленттілік.</i> Этнографизмнің әлеуметтік-демографиялық мазмұнын нысана тілдегі дәл мәдени-тарихи баламалармен сәйкестендіру.

	фразеологизмнің контекстіне (кедейлік немесе жалғыздық) байланысты аудармалар қолданылады.	Should a man <i>sit alone in sorrow</i> just because his wife has passed? One must start a family again.	Разве подобает мужчине сидеть, <i>обняв колени в одиночестве</i> из-за смерти жены? Нужно заново создавать семью.	<i>Дескриптивті-мағыналық сәйкестік.</i> Этнографизмнің визуалды бейнесі мен семантикалық мазмұнын нысана тілдің стилистикалық нормаларына сай үйлестіру.
Claude	Бұл қазақ тіліндегі эвфемизм (жұмсақ айтылым), яғни «өлді» деген сөздің орнына қолданылатын тұрақты тіркес.	- Died - Passed away - Kicked the bucket - Bit the dust - Gave up the ghost	- Умер - Скончался - Отдал концы - Приказал долго жить - Отошел в мир иной	<i>Стильдік диссонанс және семантикалық ауысу.</i> Трагедиялық этнографизмді мағыналық тұрғыдан «өлім» концептісіне ауыстырып, оны төмен стильдегі жаргондармен бұрмалау.
		Should a man <i>sit hugging his bare knees, saying "my wife died"</i> ? One must establish a household.	Разве <i>мужчина будет сидеть, обхватив свои голые колени, говоря «моя жена умерла»</i> , нужно создавать семью (дом).	<i>Денотаттық тура аудару.</i> Этнографизмнің метафоралық мағынасын («жалғыздық/ұрпақсыз қалу») тек сыртқы физикалық белгімен (жалаңаш тізе) алмастыру.

Зерттеу барысында қарастырылған этносоматикалық бірліктердің эмпирикалық жиынтығы (жиыны 500 бірлік) бойынша генеративті модельдердің (ChatGPT, Gemini, Claude) жұмысы жүйелі түрде сарапталды. Соның ішінде ұлттық-мәдени кодтың терең қабаттарын қамтитын «жүрек жалғау», «қу тізесін құшақтау», «ой бауырымдап келу», «аузына түкірту» секілді этноұғымдарды аудару және мағынасын ашу барысында генеративті модельдердің өрескел семантикалық ауытқуларға ұшырағанын жоғарыда ұсындық. Сонымен тұтас сынақ нәтижелерін аталған модельдердің жауаптары бойынша сандық-сапалық көрсеткіштері мынадай болды:

ChatGPT (GPT-4 бағдарламалық базасы): талданған этнобірліктердің жалпы көлемінің шамамен 58-62%-ында денотативті мағынаны дұрыс танығанымен, лингвомәдени астар мен контекстік прагматикада (атап айтқанда, жоғарыда аталған этноұғымдарды тәржімалауда орын алған денотаттық буквализм мен когнитивті инерция жағдайларында) тұрақты ауытқулар байқалды.

Gemini: жалпы мәтіндік логика мен ілеспе аударма бағытында тұрақты нәтиже көрсеткенімен (65-70% деңгейінде), қазақ халқының сакралды-этнографиялық кодтарын өңдеуде «прагматикалық инверсия» мен батыстық стандарттарға бейімделу үрдісі басым болды.

Claude: көркемдік мәтіндерді генерациялауда икемділік танытқанымен, этносоматикалық бірліктердің дербес мағынасын ашу барысында (50-55% деңгейінде) лингвомәдени деритуалдау қателіктеріне жиі ұрынды.

Жалпы алғанда, 500 бірліктік эмпирикалық база мен этноұғымдардың мысалында жасалған бұл сараптама үш жүйенің де ортақ артықшылығы – мәтіндік генерация мен орташа деңгейдегі мағыналық баламаны тез табу мүмкіндігі, ал негізгі шектеуі – ұлттық коннотацияларды статистикалық жиілік негізінде ғана өңдеу екенін көрсетті.

Зерттеу нысаны ретінде алынған «жүрек жалғау» соматикалық фразеологизмін талдау барысында ЖИ жүйелерінің үш түрлі когнитивтік модельге сүйенгені анықталды. ChatGPT (OpenAI) моделінің нәтижелерінде семантикалық стереотиптендіру басым. Модель «жүрек»

соматизмін қазақ мәдениетіне тән «энергия/тамақ» доменінен ажыратып, оны жаһандық «сезім/эмоция» метафоралық доменіне теліген. Ә.Кекілбаев шығармасындағы «не ішесіз?» деп анық берілген прагматикалық белгіні есепке алмастан, жүйенің «warm the heart» (жүректі жылыту) үлгісін беруі – танымдық инерцияның нақты белгісі. Көрсетілген мысал ЖИ-дің астарлы мәдениет матрицасын сезінуге қарағанда, статистикалық жағынан жиі ұшырасатын үстем мәнге арқа сүйейтінін білдіреді. Қайта *Gemini (Google)* осы соматизмнің семантикалық түйінін дөп көрсетіп, нысана тілдегі тең мәндес нұсқаларды (to grab a quick bite, заморить червячка) ұтымды қолданды. Бұл жүйенің қазақ тіліндегі «жүрек – энергия, қуат көзі» танымдық нұсқасын тани алуы ақпараттық және оқыту корпусындағы фразеологиялық қордың сенімді екендігін аңғартады. Аталған үйлесімділік жүйенің мәнмәтін мен мәдени белгіні сәйкестіре білу икемділігінің жемісі. Ең нашар көрсеткіш *Claude (Anthropic)* жүйесінде «*семантикалық демеафоралау*» мен «*денотаттық қате*» түрінде көзге түсті. Дәл осы қателіктерге зер салып жүрген А.А. Алдабергенова мен А.Қ. Жұмабекованың мақаласы да мұны қуаттайды, яғни денотаттық мағынаның анық ұсынылмауына қатысты тұжырымдарымен өзара үйлеседі (Алдабергенова, 2024: 47). Мәселен, «жүрек жалғау» тіркесін модель метафоралық мәнде емес, анатомиялық атау сияқты қабылдайды және *heart connection, cardiac junction* және орысша *соединение сердца* секілді қате денотаттармен ұсынды. Тіпті сөйлем беріліп тұрса да, жүйе сөйлем құрамындағы «қайын ағаны» да (brother-in-law) «тёща» (mother-in-law) деп қате аударып, тек семантикалық емес, *прагматикалық және гендерлік ауытқуларға* да жол берді.

Зерттеу нәтижелері ЖИ-дің қазақ фразеологизмдерін өңдеудегі «статистикалық инерциясын» көрсетті. Бұл құбылыс *Bender мен Koller-дің* «Сегізаяқ тестімен» толық сәйкес келеді (Bender, Koller, 2020): ЖИ модельдері сөздердің жиілігі мен тізбегін білгенімен, олардың артындағы қазақы дүниетанымдық мағына мен коммуникативті ниетті ажырата алмайды. Сондықтан ChatGPT-дің «жүрек жалғауды» эмоцияға телуі – мағынаны түсіну емес, тек үйренген формаларды қайта құрастыруы. «Жүрек жалғау» фразеологизмінің ЖИ тарапынан батыстық контексте (*warm the heart*) аударылуы – модельдің ағылшын тілді корпусының «мәдени империясына» бағыныштылығын көрсетеді (Cao et al., 2023). Бұл модельдің қазақ этнобірліктерін ажырата алмай, оларды америкалық нормалармен алмастыруының тікелей салдары.

Зерттеу барысында «Ой, бауырымдау» этнобірлігін интерпретациялауда ЖИ жүйелерінің барлығында дерлік «мәдени соқырлық» байқалды. *Gemini* де бұрынғы сәйкестігін жоғалтып, когнитивті қателікке бой алды. Модель бұл тіркесті «лингвомәдени деритуалдандыруға» ұшыратып, қазақ дүниетанымындағы сакралды аза тұту ритуалын күнделікті тұрмыстағы экспрессивті одағайлармен (*Oh, boy!*) шатастырған. Бұл – *Gemini*-дің этнографиялық формуланы ритуалдық контекстен бөлектеп, оны жай эмоционалды жақындық ретінде қабылдауы. Дискурстық деңгейде модель «ат қою», «шабу» сияқты маркерлердің әсерінен семантикалық ауытқуды түзетуге тырысып, контекстік репарация жасауға талпынғанымен, бастапқы лексикалық түсіндірмеде этнобірлік мәнін бұрмалап жіберді. *Gemini* жүйесінің контекст деңгейіндегі прагматикалық сигналдарды ескеруі ЖИ-дің контекст терезесі ұлғайған сайын семантикалық дәлдігі артатыны туралы зерттеулермен үндеседі (Wang, 2024). *ChatGPT* моделінде де осыған ұқсас «ритуалдық инверсия» анықталды. Модель аза тұту тіркесін *brotherly affection, camaraderie* сияқты жағымды әлеуметтік байланыс категорияларына трансформациялаған. М. Әуезов мәтініндегі қайғы сигналдарына қарамастан, модельдің «проявляющих братскую теплоту» нұсқасын ұсынуы – ЖИ-дің терең мәдени коннотацияларды өңдеудегі әлсіздігін көрсетеді. *Claude* талдауында «концептуалды трансформация» басым болып, этнографиялық сакралды мән заманауи психологиялық «bonding» терминдерімен алмастырылған. Әдеби үзіндіге балама жасау барысында *Claude* жүйесі «ой, бауырымды» жалпы бейтарап айқай мен дабыр (*oh, brothers!*) ретінде танып, стильдік алшақтыққа бой алдырған.

Ақпараттар жиынынан байқап отырғанымыздай үш жүйеде *бауыр* этнобірлігін тек туыстық мәнде ұғынуы нейро желілердің қазақ тіліндегі «қайғы әдебі» түсініктерін қабылдаудағы қауқарсыздығын байқатады. Көрсетілген «лингвомәдени аберрация» ЖИ-дің метафоралық қалыпты үстіртін сөздік деңгейде ғана игеретіндігін көрсетеді.

Зерттеу кезеңіндегі «аузына түкірту» этноұғымына қатысты талданған мәліметтер генеративті қосымшалардың метафоралық сатыға қарағанда, халықтық таным және әдеп стереотиптеріне барынша бағынатынын байқатты. Аталмыш ұғым қазақы көзқараста «қасиет дарыту, қасиет қондыру» сынды ғұрып болып саналады, ал ЖИ үлгілерінің бәрі бұны бұрыс әрі жағымсыз әлеуметтік әрекет деп түсінді. Айталық, *ChatGPT* жүйесінде «мәдени-прагматикалық деформация» көзге түсті. Үлгі бұл ғұрыптың мистикалық түсінігін «бағындырып алу, билік жүргізу, басыну, басқа шығу» мағынасына ауыстырып жіберген. Берілген мәнмәтінге зер салмай ауыз мүшесін «биліктің орталығы, бастауы» деп танып этно сипатты мәнді бағыныштылыққа балады. Көркем шығармадан алынған мәтінді аудару кезінде жүйе нақты ритуалды *placed in their care* секілді жалпылама түсінікпен бүркемелейді. *Claude* жүйесінде ең терең ауытқу – «прагматикалық инверсия» тіркелді. Модель «түкіру» әрекетін жаһандық этикалық стереотип тұрғысынан ғана бағалап, оны «қорлау, кемсіту» ретінде қабылдаған. Нәтижесінде, Т.Әлімқұловтың жағымды контекстіндегі ритуалды *kazakhs are oppressed, spat upon* (қазақтар қысым көруде, оларға түкіреді) деп аударып мағынаны бұрмалады. Бұл – *аксиологиялық инверсияның* айқын мысалы. Ал *Gemini* бұл кезеңде жоғары лингвомәдени адаптация көрсетті. Модель ритуалдың магиялық мәнін сақтап, нысана тілдегі *to pass the torch* сияқты идиомалық баламаларды ұсынған. Көркем мәтінді аударуда *spit in the infant's mouth for good luck* (сәбидің аузына сәттілік үшін түкіру) деген түсіндірмені қосуы – экспликацияланған сәйкестіктің сәтті үлгісі. Бұл – модельдің мәдени шок тудыруы мүмкін элементті прагматикалық бейімдеу қабілеті.

ЖИ жүйелерінің қазақ тіліндегі этнографиялық мазмұнды «қамқорлық», «бағындыру» немесе «қорлау» деп трансформациялауы ауызша дискурста тән сызықтық емес (*non-linearity*) сипаттың машиналық интеллект тарапынан дұрыс бағаланбауынан. *X. Wang және C. Fantinuoli* зерттеулеріне сәйкес, автоматты метрикалар мен тілдік модельдер контексттік нюанстарды адамша толық қабылдай алмайды (Wang, 2024). Ал бұны зерттеуіміздегі ЖИ үлгілерінің сакралды дәстүрлерді халықтық-тұрмыстық қабатқа қарай құнсыздандырып тастауынан көруге болады. ЖИ үлгілерінің «аузына түкірту» этноұғымын түсінуіндегі шиеленістер әмбебап этикалық нормалар мен тар шеңбердегі этномәдени кодтар ортасындағы теңдікті ұстап тұра алмайтынын байқатады. *Gemini*-ден басқасы ғұрыпты «бағындыру» не «қорлау» ретінде қабылдауы ЖИ-дің қазақ мәдениетіне тән мистикалық шындықты түсінуге деген ұмтылыс емес, танымдық шектеу болып саналады.

Зерттеудің нысаны деп алынған «қу тізесін құшақтау» тіркесіне сараптама жасауда ЖИ үлгілерінің метафоралық образды пайымдауында концептілік диссонанс анық байқалды. Қазақ тіліндегі аталмыш ұғым әлеуметтік-демографиялық жағдайды, яғни жалғыздық, ұрпақсыз қалу деген мәнді білдірсе, жасанды жүйелер бұны материалдық жетіспеушілікпен шатастырады. Анығырақ тоқталсақ, *ChatGPT* бойынша *концептуалды алмасу* тіркелді. ЖИ қу лексемасын «бос» деп түсініп, мағынаны «кедейлік» реңкіне бағыттады. Т. Ахтанов үзіндісіндегі берілген мәнмәтінге қарамастан, жүйенің «материалдық кедейлік» үлгісінен ажырамай тұрып қалуы семантикалық инерцияны байқатады. Баламадағы «empty knees» (бос, қуыс тізе) тіркесін мәні жоқ, жасанды құрылым ретінде бағалауға болады. *Claude* үлгісінде «*стильдік диссонанс*» және «*денотаттық буквализм*» бірге байқалды. Жүйе бұл этноұғымды «өлім» концептісімен байланыстырып, бұны *kicked the bucket* сынды теріс әрі жағымсыз стильдегі жаргондармен аударма жасап беріп тұр. Бұны ЖИ-дің қасіретті пафосты сезінбей стилистикалық жағынан сәйкеспейтін эквиваленттермен алмастыру жағдайы деп танимыз. Әдеби мәтінге аударма жасаудағы *hugging his bare knees* нұсқасы жүйенің фразеологиялық ұғымды жай ғана физикалық әрекет деп қабылдайтынын дәлелдейді. *Gemini* үлгісі осы бөлімде сауатты функционалдық эквиваленттілік байқата алды. Себебі жүйе этноұғымның әлеуметтік-демографиялық сипатын дәл анықтап қана қоймай, *high and dry, alone in sorrow* сынды баламаларды ұтымды таңдады. Ал орыс тіліне аударудағы *остаться бобылем* баламасын мәдени-тарихи контекст сақталған көрсеткіш деп айта аламыз.

Сонымен, «қу тізесін құшақтау» ұғымына қатысты сараптама нәтижесі ЖИ үлгілерінің қазақ тіліндегі соматикалық белгілерді дешифрлеудің түрлі сатысын байқатты. *ChatGPT* мен *Claude*-нің «тізе» мен «қу» сөздерін жай ғана физикалық негіз деп түсінуі машиналық баламаның

абстракттілі әрі образды ойлауға ілесе алмайтындығын білдіреді. Ал Gemini-дің өзге модельдерден артықшылығы лингвистикалық базасының *мағына-контекст* үйлесімділігі бойынша жақсы дамып келе жатқанын көрсетеді.

Қорытынды

Жүргізілген зерттеу нәтижелері қазіргі ЖИ модельдердің (ChatGPT, Gemini, Claude) қазақ тіліндегі соматикалық этнографизмдерді өңдеудегі когнитивті шектеулері мен аудармадағы семантикалық ауытқуларын айқындады. Зерттеу барысында алынған тұжырымдарды төмендегідей қорытындылауға болады:

Когнитивтік модельдеудегі ашықтық: ЖИ жүйелерінің көпшілігінде (әсіресе ChatGPT пен Claude) «когнитивті инерция» құбылысы анықталды. Көркем мәтін ішіндегі нақты прагматикалық сигналдарға («не ішесіз?», «әйелім өлді») қарамастан, модельдер этнобірліктерді жаһандық статистикалық басым мағыналармен (эмоция, физикалық әрекет) интерпреттеуді жалғастырды. Бұл *Bender мен Koller-дің* (Bender, Koller, 2020) ЖИ-дің таңба мен шынайы мағына арасындағы байланысты өздігінен орната алмайтындығы туралы гипотезасын қазақ тілі материалында көрсетеді.

Семантикалық ауытқулардың сипаты: Зерттеу барысында анықталған *прагматикалық инверсия* («аузына түкіртуді» қорлау деп түсіну), *лингвомәдени деритуализация* («бауырымдауды» жай одағайға теңеу) және *денотативті буквализм* («қу тізесін құшақтауды» физикалық әрекет деп тану) құбылыстары модельдердің ағылшын тілді корпустың мәдени империясына (Cao et al., 2023) бағыныштылығын көрсетеді. Қазақы этнографиялық кодтар ЖИ тарапынан батыстық этикалық нормалар сүзгісінен өтіп, өзінің сакралды-мәдени мазмұнын жоғалтты.

Жүйелердің функционалдық сатысы бойынша: Gemini моделі басқа жүйелермен салыстырғанда жоғары тілдік әрі мәдени бейімдеу мен жұмсалымдық тепе-теңдік көрсетті. Бұл аталмыш үлгінің қазақ тіліндегі «контекст үлгісін» оңтайлы қолдана алатынын және оқыту дерекқорындағы фразеологиялық бөлімнің сапалы екенін байқатады. Ал Claude моделіндегі мағыналық ауытқулар метафоралық ойлау сатысының әлі де болса төмен әрі сенімсіз екенін растайды.

Лингвомәдени аберрация және цифрлық тұтастық негізінде: ЖИ тарапынан жіберілетін қателіктер қазақ тілінің цифрлық кеңістіктегі семантикалық тұтастығына нұқсан келтіреді. Модельдердің қазақ этнографизмдерін америкалық стандарттармен алмастыруы ұлттық тілдік бейненің бұрмаланып кетуіне әкеп соғады.

Түйіндей айтқанда, қазіргі ЖИ жүйелері қазақ тілінің беткі лексикалық қабатын өңдеуде ілгерілеушілік көрсеткенімен, «қайғы мәдениеті», «магиялық реализм» және «соматикалық кодтар» сияқты терең этнолингвистикалық қабаттарды декодтауда әлі де жетілдіре түсуді қажет етеді. Зерттеу нәтижелерінен зерттеудің болашағы мен практикалық маңызына қатысты мынадай ұсыныстар туындады: алынған деректер ЖИ модельдерін дамытуда тек статистикалық мәліметтерді емес, терең этнографиялық контекстермен нығайтылған, мамандандырылған деректер қорын пайдалану қажеттігін негіздейді. Сондай-ақ алынған нәтижелерді кең көлемді корпустық деректер мен модельдердің жанартылған нұсқаларында қайта тексеру – болашақ зерттеулердің негізгі бағыты. Бұл бағыттағы жұмыстар қазақ тілінің этнографиялық базасын модельдерге тереңірек енгізу және арнайы баптау (fine-tuning) арқылы лингвомәдени адаптациясын арттыруға мүмкіндік береді.

Әдебиеттер

Абдрахманова М.Б. «Көз» концепті: лингвомәдени және танымдық парадигмасы: филол. ғыл. канд. дис. – Алматы, 2008.

Алдабергенова А.А., Жұмабекова А.Қ. Жаңа аударма мәтіндерінің баламалылық деңгейі // Торайғыров университетінің Хабаршысы. Филологиялық серия. – 2024. – №3. – 44-52-бб.

Барменқұлова А.С., Жаңабекова А.Ә., Манкеева Ж.А. Этномәдени бірліктердің корпуста берілу жолдары // *Тілтаным*, 2023. №1 (56). – 56-66-бб.

Ескермесова Г.А., Ермекова Т.Н., Шойбекова Ғ.Б. Этномәдени бірліктердің жасанды интеллект арқылы аударылуы: машиналық аудармалар және оның сапасын бағалау // Еуразия гуманитарлық институтының

хабаршысы. – 2026. – №2. – 67–74-бб. DOI: <https://doi.org/10.55808/1999-4214.2026-2.06>

Жанабекова А.А., Усенов М., Тлегенова Г.Б. Создание больших языковых моделей в казахском языке: казахстанский опыт и глобальные тренды // *Tiltanyam*, 2026. №1 (101). – С. 135–149. DOI: <https://doi.org/10.55491/2411-6076-2026-1-135-149>

Имангазина М. Аудармадағы коммуникант эмоциясы: түр-түс атауларының қызметі мен семантикасы // Вестник евразийского гуманитарного института. – 2024. – №3. – С. 41-52.

Исаева Г.С., Өтебаева Э.Ә., Серікқызы М. Лингвомәдени ішкорпусты әзірлеудегі мәдени-семантикалық белгіленім // *Тіланым*, 2023. №3 (135). – 135-143-бб.

Ислам А. Медиалингвистика негіздері және медиа аударма мәселелері // Қарағанды университетінің хабаршысы. Филология сериясы. – 2025. – №2 (118). – 151-160-бб.

Қазақ әдеби тілінің сөздігі. Он бес томдық. 6-том / Құраст.: Ж. Қоңыратбаева, Ғ. Қалиев, Қ. Есенова және т.б. – Алматы, 2011. – 752 б.

Қазақ әдеби тілінің сөздігі. Он бес томдық. 10-том / Құраст.: М. Малбақов, Н. Оңғарбаева, А. Үдербаев және т.б. – Алматы, 2011. – 752 б.

Қазақтың этнографиялық категориялар, ұғымдар мен атауларының дәстүрлі жүйесі. Энциклопедия. – Т. 1: А-Д / Ғылыми ред. Н. Әлімбаев. – Алматы: DPS, 2011. – 735 б.

Қайдар Ә. Қазақтар ана тілі әлемінде (этнолингвистикалық сөздік). – Т. 1: Адам. – Алматы: Дайк-Пресс, 2009. – 784 б.

Манкеева Ж. Қазақ тіл білімінің мәселелері. – Алматы: «Абзал-Ай» баспасы, 2014. – 640 б.

Нұржанова А. Дүниенің қарабайыр бейнесі: фразеологиялық фрагменттің соматикалық коды: филол. ғыл. канд. ... дис. – Астана, 2010.

Раманова А.А. Қазақ тіліндегі соматизмдердің лексикалық-семантикалық өрісі // Батыс Қазақстан инновациялық-технологиялық университетінің хабаршысы. – 2023. – Т. 68. №1. – 101-110-бб.

Смағұлова Г. Қазақ фразеологиясы лингвистикалық парадигмаларда: монография. – Алматы: «Елтаным баспасы», 2020. – 256 б.

Сыздық Р. Ғылыми таным үзіктері: Мақалалар жинағы. 1-кітап. / Рабиға Сыздық. – Алматы: «Сардар», 2025. Қазақ және орыс тілдерінде.

Ташауов Ә.С. «Кодекс Куманикус» тіліндегі соматикалық атаулар және олардың лексика-семантикалық топтары: филол. ғыл. канд. ... дис. — Алматы, 1995. – 161 б.

Уәли Н. Қазақ сөз мәдениетінің теориялық негіздері. Монография. – Алматы: «Қазақ тілі» баспасы, 2021. – 424 б.

Уызбаева Б.К. Қазақ тіліндегі соматикалық етістік фразеологизмдердің этнолингвистикалық сипаты: филол. ғыл. канд. ... дис. – Алматы, 1994.

Alimzhanova, Sh.A. The role of somatisms in polysemantic phraseological units: a cognitive and cultural feature // *International Journal of Literature and Languages*. 2025. Vol. 5. Issue 03. Pp. 74-77.

Bender, E.M., Koller, A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data // *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020. Pp. 5185–5198.

Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., Hershovich, D. Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study // *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2023. Pp. 53–67.

Kadyrbek, N., Tuimebayev, Z., Mansurova, M., Viegas, V. The Development of Small-Scale Language Models for Low-Resource Languages, with a Focus on Kazakh and Direct Preference Optimization // *Information*. 2025. Vol. 16(1). Art. No. 85.

Karyukin, V., Rakhimova, D., Karibayeva, A., et al. The neural machine translation models for the low-resource Kazakh–English language pair // *Scientific Reports*. 2024. Vol. 14. Art. No. 7166.

Marcus, G. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence // *arXiv preprint*. 2020. arXiv:2002.06177. Pp. 1-45.

Maxutov, A., Myrzakhmet, A., Braslavski, P. Do LLMs Speak Kazakh? A Pilot Evaluation of Seven Models // *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2024. Pp. 112–128.

Tekgürler, M. LLMs for Translation: Historical, Low-Resourced Languages and Contemporary AI Models // *Proceedings of LaTeCH-CLfL 2025*. Pp. 227–237.

Togmanov, M., Mukhituly, N., Turmakhan, D., et al. KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan // *MBZUAI Technical Report*. 2025. Pp. 1-22.

Toleu, A., Tolegen, G., Ualiyeva, I. Fine-Tuning Large Language Models for Kazakh Text Simplification // *Applied Sciences*. 2024. Vol. 14(5). Art. No. 2100.

Wang, X., Fantinuoli, C. Exploring the Correlation between Human and Machine Evaluation of Simultaneous Speech Translation // *Proceedings of the 25th Annual conference of the European Association for Machine Translation (Volume 1)*. 2024. Pp. 327–336.

References

Abdrahmanova, M.B (2008) “Koz” koncepti: lingvomadeni zhane tanymdyq paradigmasi: filol. gyl. kand. dis. Almaty. [Abdrakhmanova, M.B (2008) The “Eye” concept: linguocultural and cognitive paradigm: Candidate of Philological

Sciences dissertation. Almaty.] (in Kazakh)

Aldabergenova, A.A., Zhumabekova, A.K. (2024) Zhanga audarma matinderining balamalylyq denggeji. Toraigyrovo universitetining Habarshysy. Filologijalyq serija. No. 3. 44-52-bb. DOI 10.48081/HUHW2781. [Aldabergenova, A.A., Zhumabekova, A.K. (2024) The level of equivalence of new translation texts. Bulletin of Toraigyrovo University Philological series. No. 3. Pp. 44-52.] (in Kazakh)

Alimzhanova, Sh.A. (2025) The role of somatisms in polysemantic phraseological units: a cognitive and cultural feature. International Journal of Literature and Languages 5(03). Pp. 74-77. DOI 10.37547/ijll/Volume05Issue03-19. (in English)

Barmenkulova, A.S., Zhagnabekova, A.A., Mankeeva, Zh.A. (2023) Jetnomadeni birlikterding korpusta berilu zholdary. *Tiltanyim*. No. 1 (56). 56-66-bb. DOI 10.55491/2411-6076-2023-1-56-66. [Barmenkulova, A.S., Zhanabekova, A.A., Mankeyeva, Zh.A. (2023) Ways of transmission of ethnocultural units in the corpus. *Tiltanyim*. No. 1 (56). Pp. 56-66.] (in Kazakh)

Bender, E.M., Koller, A. (2020) Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 5185–5198. (in English)

Cao, Y., Zhou, L., Lee, S., Cabello, L., Chen, M., Hershcovich, D. (2023) Assessing Cross-Cultural Alignment between ChatGPT and Human Societies: An Empirical Study. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP). Pp. 53–67. (in English)

Esfermesova, G.A., Ermekova, T.N., Shojbekova, G.B. (2026) Jetnomadeni birlikterding zhasandy intellekt arqyly audaryluy: mashinalyq audarmalar zhane onyng sapasyn bagalau. Evrazija gumanitarylyq institutynyng Habarbarysy. No. 2. 67–74-bb. DOI <https://doi.org/10.55808/1999-4214.2026-2.06> [Yesfermesova, G.A., Yermekova, T.N., Shojbekova, G.B. (2026) Translation of ethnocultural units through artificial intelligence: machine translation and evaluation of its quality. Bulletin of the Eurasian Humanities Institute. No. 2. Pp. 67–74.] (in Kazakh)

Imangazina, M. (2024) Audarmadagi kommunikant jemocijasy: tur-tus ataularynyng qyzmeti men semantikasy. Vestnik Evrazijskogo gumanitarnogo instituta. No. 3. 41-52-bb. [Imangazina, M. (2024) Communicant emotion in translation: functions and semantics of color terms. Bulletin of the Eurasian Humanities Institute. No. 3. Pp. 41-52.] (in Kazakh)

Isaeva, G.S., Otebaeva, Je.A., Serikqyzy, M. (2023) Lingvomadeni ishkorpusty azirleudegi madeni-semantikalyq belgilenim. *Tiltanyim*. No. 3 (135). 135-143-bb. DOI 10.55491/2411-6076-2023-3-135-143. [Issayeva, G.S., Utebayeva, E.A., Serikqyzy, M. (2023) Cultural and semantic markup in the development of the linguocultural subcorpus. *Tiltanyim*. No. 3 (135). Pp. 135-143.] (in Kazakh)

Islam, A. (2025) Medialingvistika negizderi zhane media audarma maseleleri. Qaragandy universitetining Habarshysy. Filologiya serijasy. No. 2 (118). 151–160-bb. [Islam, A. (2025) Fundamentals of medialinguistics and issues of media translation. Bulletin of the Karaganda University Philology series. No. 2 (118). Pp. 151–160.] (in Kazakh)

Kadyrbek, N., Tuimebayev, Zh., Mansurova, M., Viegas, V. (2025) The Development of Small-Scale Language Models for Low-Resource Languages, with a Focus on Kazakh and Direct Preference Optimization. Information. 16(1) 85. DOI 10.3390/info16010085. (in English)

Karyukin, V., Rakhimova, D., Karibayeva, A. et al (2024) The neural machine translation models for the low-resource Kazakh–English language pair. Scientific Reports. 14 7166. DOI 10.1038/s41598-024-57854-x. (in English)

Mankeeva, Zh. (2014) Qazaq til bilimining maseleleri. Almaty: “Abzal-Ai” baspasy. 640 b. [Mankeyeva, Zh. (2014) Issues of Kazakh linguistics. Almaty: “Abzal-Ai” publishing house. 640 p.] (in Kazakh)

Marcus, G. (2020) The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. arXiv preprint arXiv:2002.06177. Pp. 1-45. (in English)

Maxutov, A., Myrzakhmet, A., Braslavski, P. (2024) Do LLMs Speak Kazakh? A Pilot Evaluation of Seven Models. Proceedings of EMNLP. Pp. 112–128. (in English)

Nurzhanova, A. (2010) Duniening qarabajyr bejnesi: frazeologijalyq fragmentting somatikalyq kody: kand. dis. Astana. [Nurzhanova, A. (2010) Primitive picture of the world: somatic code of the phraseological fragment: Candidate dissertation. Astana.] (in Kazakh)

Qajdar, A. (2009) Qazaqtar ana tili aleminde (jetnolingvistikalyq sozdik). T. 1: Adam. Almaty Dajk-Press. 784 b. [Kaidar, A. (2009) Kazakhs in the world of their native language (ethnolinguistic dictionary). Vol 1: Man. Almaty: Dyke press. 784 p.] (in Kazakh)

Qazaq adebi tilining sozdigi. On bes tomdyq. 6 tom (2011a) Almaty, 752 b. [Dictionary of the Kazakh literary language. In fifteen volumes. Vol. 6. (2011a) Almaty, 752 p.] (in Kazakh)

Qazaq adebi tilining sozdigi. On bes tomdyq. 10 tom. (2011b) Almaty, 752 b. [Dictionary of the Kazakh literary language. In fifteen volumes. Vol. 10 (2011b) Almaty, 752 p.] (in Kazakh)

Qazaqtyn jetnografijalyq kategorijalar, uymdar men ataularynyng dasturli zhujesi. T. 1. A-D (2011) Almaty: DPS. 735 b. [Traditional system of Kazakh ethnographic categories concepts and names. Vol. 1. A-D (2011) Almaty: DPS. 735 p.] (in Kazakh)

Ramanova, A.A. (2023) Qazaq tilindegi somatizmderding leksikalyq-semantikalyq orisi. Batys Qazaqstan innovacijalyq-tehnologijalyq universitetining Habarshysy. T. 68. No. 1. 101-110-bb. DOI 10.48371/PHILS.2023.23.18.011. [Ramanova, A.A. (2023) Lexical and semantic field of somatisms in the Kazakh language. Bulletin of the West Kazakhstan Innovation and Technology University. Vol. 68. No. 1. Pp. 101-110.] (in Kazakh)

Smagulova, G. (2020) Qazaq frazeologijasy lingvistikalyq paradigmalarda: monografija. Almaty: “Eltanyim baspasy”. 256 b. [Smagulova, G. (2020) Kazakh phraseology in linguistic paradigms: monograph. Almaty: “Eltanyim publishing

house”. 256 p.] (in Kazakh)

Syzdyq, R. (2025) Gylimi tanym uzikteri: Maqalalar zhinagy. 1-kitap. Almaty: “Sardar”. Qazaq zhane orys tilderinde. [Syzdyk, R. (2025) Fragments of scientific knowledge: Collection of articles. Book 1. Almaty: “Sardar”. In Kazakh and Russian languages.] (in Kazakh)

Tashauov, A.S. (1995) “Kodeks Kumanikus” tilindegi somatikalyq ataular zhane olardying leksika-semantikalyq toptary: filol. gyl. kand. dis. Almaty. 161 b. [Tashauov, A.S. (1995) Somatic names in “Codex Cumanicus” and their lexical-semantic groups: Candidate of Philological Sciences dissertation. Almaty. 161 p.] (in Kazakh)

Tekgürler, M. (2025) LLMs for Translation: Historical, Low-Resourced Languages and Contemporary AI Models. Proceedings of LaTeCH-CLfL. Pp. 227–237. (in English)

Togmanov, M., Mukhituly, N., Turmakhan, D. et al (2025) KazMMLU: Evaluating Language Models on Kazakh, Russian, and Regional Knowledge of Kazakhstan. MBZUAI Technical Report. Pp. 1-22. (in English)

Toleu, A., Tolegen, G., Ualiyeva, I. (2024) Fine-Tuning Large Language Models for Kazakh Text Simplification // Applied Sciences. Vol. 14(5). Art. No. 2100. (in English)

Uali, N. (2021) Qazaq soz madenietining teorijalyq negizderi. Monografija. Almaty: Qazaq tili baspasy. 424 b. [Uali, N. (2021) Theoretical foundations of Kazakh speech culture. Monograph. Almaty: Qazaq tili publishing house. 424 p.] (in Kazakh)

Uyzaeva, B.K. (1994) Qazaq tilindegi somatikalyq etistik frazeologizmderning jetnolingvistikalyq sipaty: filol. gyl. kand. dis. Almaty. [Uyzaeva, B.K. (1994) Ethnolinguistic nature of somatic verbal phraseologisms in the Kazakh language: Candidate of Philological Sciences dissertation. Almaty.] (in Kazakh)

Wang, X., Fantinuoli, C. (2024) Exploring the Correlation between Human and Machine Evaluation of Simultaneous Speech Translation // Proceedings of the 25th Annual conference of the European Association for Machine Translation. Vol. 1. Pp. 327–336. (in English)

Zhanabekova, A.A., Usenov, M., Tlegenova, G.B. (2026) Sozdanie bol'shih jazykovykh modelej v kazahstanskom jazyke: kazahstanskij opyt i global'nye trendy. *Tiltanyim*. No. 1. S. 135–149. DOI <https://doi.org/10.55491/2411-6076-2026-1-135-149>. [Zhanabekova, A.A., Usenov, M., Tlegenova, G.B. (2026) Creation of large language models in the Kazakh language: Kazakhstani experience and global trends. *Tiltanyim*. No. 1. Pp. 135–149.] (in Russian)

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

Редакцияға түсті / Поступила в редакцию / Entered the editorial office: 19.06.2026.

Жариялауға қабылданды / Принята к публикации / Accepted for publication: 22.09.2026.

Интернетте қолжетімді / Доступно онлайн / Available online: 30.09.2026.

© Қамыза Б.Қ., Еркемова Т.Н., Ескермесова Г.А., 2026

