ISSN 2788-7979 (online)

ӘДЕБИЕТТАНУ ЖӘНЕ ФОЛЬКЛОРТАНУ

ЛИТЕРАТУРОВЕДЕНИЕ И ФОЛЬКЛОРИСТИКА

LITERARY STUDIES AND FOLKLORISTICS

ҒТАМР 16.31.21 DOI: 10.59102/kufil/2026/iss3pp243-256

Е.А. Амангельдин[1], Ә.С. Мусаев[1], Д.А. Алкебаева[1]
[1]Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, 050000, Қазақстан

ЖАСАНДЫ ИНТЕЛЛЕКТ ЖӘНЕ ҚАЗАҚ ТІЛІНІҢ ҰЛТТЫҚ-МӘДЕНИ БОЛМЫСЫ: СЕМАНТИКАЛЫҚ ЖӘНЕ ПРАГМАТИКАЛЫҚ ИНТЕРПРЕТАЦИЯ МӘСЕЛЕЛЕРІ

Қазіргі цифрлық даму үдерісі жағдайында тілдің табиғаты мен қызметі жаңа технологиялық ортада қайта қарастырылуда. Әсіресе жасанды интеллект жүйелерінің (ЖИ) қарқынды өркендеуі ұлттық тілдердің цифрлық әлеуетін бағалауды өзекті ғылыми мәселеге айналдырып отыр. Осыған байланысты зерттеу жұмысы қазақ тіліндегі этномәдени лексиканың ЖИ модельдері арқылы игерілу әлеуетін анықтауға бағытталған. Қазақ тілінің семантикалық, прагматикалық және стилистикалық ерекшеліктерін ЖИ жүйелері арқылы талдап, олардың ұлттық-мәдени кодты тану қабілетін салыстырмалы тұрғыда бағалау басты мақсат ретінде алынды. Негізгі идея – тілдің тек құрылымдық жүйе емес, ұлттық дүниетанымды бейнелейтін когнитивтік феномен екенін ескере отырып, ЖИ-дің оны меңгеру құзыреттілігін анықтау.

Зерттеудің ғылыми және практикалық маңызы қазақ тілінің цифрлық кеңістіктегі даму қадамдары айқындаумен және ЖИ модельдерін ұлттық тілге бейімдеу мәселелерін негіздеумен сипатталады. Әдіснамалық тұрғыдан жұмыс контекстік, когнитивті-семантикалық және прагматикалық талдау әдістеріне сүйеніп, эксперименттік тәсіл арқылы жүзеге асырылды.

Эксперимент қорытындысы ЖИ модельдерінің қазақ тілінің ұлттық кодын меңгеру деңгейі орташа есеппен 57%-ды құрайтынын көрсетті. Модельдер арасында Grok (73%) ең жоғары дәлдік танытса, DeepSeek (40%) ең төменгі көрсеткішке ие болды. Тілдік модельдердің грамматикалық деңгейдегі сауаттылығына қарамастан, прагматикалық реңктерді ажыратуда және этнографиялық мағынаны интерпретациялауда жүйелі қателіктер мен «алгоритмдік галлюцинациялар» анықталды.

Зерттеудің құндылығы – қазақ тілінің ұлттық-мәдени ерекшеліктерін цифрлық технологиялар контексінде кешенді түрде қарастыруында. Жұмыс нәтижелері ЖИ жүйелерін қазақ тіліне бейімдеу, тілдік деректер қорын кеңейту және білім беру үдерісінде тиімді қолдану бағытында практикалық маңызға ие.

Кілт сөздер: жасанды интеллект, үлкен тілдік модельдер, этномәдени лексика, ұлттық код, когнитивті лингвистика, прагматика, семантика, стилистика.

НЕГІЗГІ ЕРЕЖЕЛЕР

Қазіргі цифрлық дәуір адамзаттың жаңа өмір сүру тәртібі мен заңдылықтарын қалыптастыра отырып, барлық заман келбетінің әлеуметтік және мәдени тұтастығын түбегейлі өзгертіп жатыр. Цифрлық үдеріс технологиялық өркендеудің маңызды қозғалтқышы ретінде ғана емес, лингвистиканың да тұрақталған ішкі жүйесіне төңкеріс әкелетін күшке ие болып, тілдің болашағына айтулы ықпалын тигізуде.

ISSN 2788-7979 (online)

Мемлекет басшысы Қасым-Жомарт Тоқаевтың шешімімен 2026 жылдың «Цифрландыру және жасанды интеллект жылы» болып бекітілуі отандық ғылым үшін жаңа технологиялық дәуірдің бастауын айқындады. Бұл басымдық ғалымдар қауымына зерттеу процестерін цифрлық трансформациялау және жасанды интеллект мүмкіндіктерін ғылыми интеграциялау бойынша ауқымды міндеттер жүктейді. Еліміздің бар ғылым салалары аталған цифрландыру әрекетіне баса назар аударып, даму жылдамдығын одан әрі күшейту мақсатында жаңа қадамдар жасайтыны сөзсіз [1].

Жасанды интеллект (ЖИ) соңғы бес жылдың ішінде ірі инновациялық жаңалық ретінде дүниежүзіне танылып үлгерді. Арнайы алгоритмдік жүйелер бұрын компьютердің ойлауға қабілетсіз есептік көрсеткіштерін жандандырып, олардың тек аса күрделі математикалық операцияларын орындайтын ғана емес, адамның танымдық және тілдік қабілеттерін терең модельдеуге де жол ашты. Бұл эволюция машиналардың тек сандық мәліметтермен шектеліп қалмай, табиғи тілді өңдеу (NLP) кеңістігіне өтуіне түрткі болды. Үлкен тілдік модельдердің (LLM) қарқынды дамуы ЖИ-ге мәтіндерді жай ғана генерациялап қоймай, олардың семантикалық мағыналарын ажыратуға, прагматикалық контексті түсінуге және стилистикалық ерекшеліктерді сақтауға мүмкіндік беріп отыр.

Алайда, екінші жағынан, бұл жаһандық алгоритмдік жетістіктер әлемнің барлық тілдері үшін бірдей тиімділік көрсете бермейді. Әсіресе, грамматикалық табиғаты мен морфологиялық құрылымы өзгеше, жалғамалы тілдердің (соның ішінде қазақ тілінің) цифрлық әлеуетін толық ашу – күрделі міндет. ЖИ жүйелерінің ұлттық тілдің бірегей лингвистикалық бірліктерін қаншалықты дәл өңдейтінін бағалау және табиғи тіл мен машиналық логиканың өзара әрекеттесуін қарастыру қазіргі филология мен есептеуіш ғылымдардың тоғысындағы ең маңызды зерттеу нысандарының біріне айналды.

Қазақтың басты байлығы – оның тілі. Бұл – сан ғасырлар бойы, атадан балаға аманат болған асыл мұра. Тілдегі әрбір құбылыс өздігінен немесе кездейсоқ пайда болған жоқ. Ол халықтың тарихымен, тағдырымен тығыз байланыста дамыды.

Қазақ тіліндегі ұлттық нақышқа бай сөздер мен тіркестер ұлттық болмысымызбен бірге өріліп, рухани тамырымыздан нәр алды. Әсіресе, қазақтың біртуар перзенттері: жыраулар, бишешендер, Абай, Шәкәрім, Махамбет, Мұхтар сынды ақын-жазушыларымыз әдеби тіліміздің берік іргетасын қалап, ана тіліміздің мәртебесін асқақтатып, оның көркемдік қуатын әлемдік деңгейге көтерді.

Тілдің ұлттық қазынасы мен байлығын жіті зерттей отырып, өзге халықта жоқ теңдессіз ерекшеліктерін алға шығаратын бағыт – этнолингвистика саласы. Бұл бағытта қазақ тіл білімінің ұлттық сана, таным ұғымдарының төңірегінде құнды қағидаларымен толтыра білген Ә. Қайдар, Р. Сыздық, Е. Жанпейісов, Н. Уәли, Ж. Манкеева, Г. Смағұлова, Б. Момынова сынды көрнекті ғалымдар мол еңбектер қалдырды. Зерттеушілердің басты назарында тілдегі сөздердің тек этимологиялық қыр-сыры емес, олардың халықтың танымдық, философиялық білімімен ұштасуы болды.

Тіл біліміндегі антропоцентристік парадигманың іргетасы саналатын Э. Сепир мен Б. Л. Уорфтың «лингвистикалық салыстырмалылық» гипотезасына сәйкес, тіл – ақпарат тасымалдаушы механикалық құрал ғана емес, ол ұлттың ойлау қабілеті мен дүниетанымын қалыптастыратын бірегей когнитивтік жүйе. Бұл тұжырымдама әрбір этностың қоршаған ортаны объективті түрде емес, ғасырлар бойы қалыптасқан өз ана тілінің лексика-грамматикалық заңдылықтары мен ұлттық-мәдени бірліктерінің призмасы арқылы қабылдайтынын дәлелдейді [2, 210]. Осы орайда, ұлттық болмысты таңбалайтын семантикалық бірліктердің табиғатын жалаң таңбалар жиынтығы ретінде емес, халықтың тарихи жады, діли мен психологиясы біте қайнасқан, басқа ешбір жүйеге дәл сол танымдық қуатымен көшірілмейтін тірі әрі дербес құбылыс ретінде бағалау – ғылыми зерттеудің теориялық негізін айқындай түседі.

Академик Ә. Қайдар: «Шынында да, Адам баласы адамзат дүниесін, бүгінгі материалдық болмысын тек ақыл-парасаты арқылы ғана танып-біліп, жан-жақты дамытып отыр. Демек, Адамның ең басты, ең қымбат, ең абзал, ең құнды қасиеті – оның Адамдық қасиеті» дей келе,

ISSN 2788-7979 (online)

адамның күллі әлемді игеруі оның басты құралы «ақылды» пайдалануы арқылы қолжетімді болғанын алға тартады [3, 12]. Яғни, қазіргі зор жетістіктердің негізінде тікелей адамның күрделі когнитивті қабілеті мен зиялы қасиеті жатыр. Сонымен қатар, академик ұлттық тілдік бірліктер ғасырлар бойы қалыптасқан халықтың өмірлік тәжірибесінен, ал оның негізі ұйытқысы саналы адамның қоршаған ортамен етене араласуынан, табиғат пен қоғамның ықпалынан туындайтынын дәлелдейді. Ғалымның этнолингвистикалық сөздіктері қазақ тілінің лексикалық байлығы бір арнада шоғырланған артефактілері іспетті бағаланып, көптеген зерттеулерге мұрындық болады.

Қазақ тіл біліміндегі этнолингвистикалық бағытты сөздің тарихи семантикасымен ұштастыруда академик Р. Сыздықтың ғылыми мұрасының орны ерекше. Академиктің тарихи-этнолингвистикалық зерттеулерінің басты мұраты мән-мағынасы уақыт өте келе күңгірттеген, жұрттың жадында пассивтенген тілдік бірліктердің ішкі семантикасын тірілтуге күш жұмсау болды. XV – XIX ғасырлардағы ақын-жыраулар поэзиясы мен батырлар жырларындағы оқырманға мүлдем таңсық болып қалған сөздермен қатар қазіргі таңда көпшілік белсенді қолданып жүрген сөздердің ертеректе астарында тіпті өзге мағына жүктелгендігін растайды [4, 5]. Ғалымның бұл ғылыми тұжырымдары тілдің тірі организм ретінде үнемі өзгеріске бейім құбылыс екенін дәлелдей отырып, оның «жанды» антропоөзекті табиғатын назардан тыс қалдырмау керектігін айқындайды.

Дәстүрлі лингвистиканың өкілдері ұзақ жылдар бойы тілдің барлық аспектілерін дамытуға күш жұмсады. Олардың тілге қатысты диахронды, синхронды зерттеу еңбектері бүгінгі таңда жас ізденушілер үшін теориялық және практикалық маңызы бар құнды деректерге айналды. Әсіресе, заманауи тілдің жаңалықтары мен мәселелеріне әділ баға беру мақсатында бұрынғы ғылыми парадигмалар мен бүгінгі инновациялық үрдістерді өзара сабақтастыра қарастыру қажеттілігі туындайды. Мұндай сабақтастық тілдің құрылымдық-семантикалық болмысын, ұлттық ерекшеліктерін цифрлық кеңістікте бұрмалаусыз сақтап қалуға жол ашады.

КІРІСПЕ

Қазақ тілінің лексикалық қоры ғасырлар бойы атадан-балаға мұра болып жеткен, әрбір сөз бен оралымда халқымыздың өмір тәжірибесі, дәстүрі, тұрмыс-салттары мен философиялық ойлары сақталған. Тілдегі ұғымдар бір-бірімен нәзік байланыста, әртүрлі жағдайларда жаңа мәнге ие болып, сөйлеушінің ойлау және сезім өрісін дәл жеткізуге мүмкіндік береді. Тілдің фразеологиялық байлығы мен стилистикалық әртүрлілігі, сонымен қатар синонимдік және омонимдік ресурстары оның семантикалық әлеуетін кең етіп қана қоймай, қайталанбас әрі жатқа танылмас сипатқа айналдырады. Бұл ерекшеліктер тілдің адами факторларға тәуелді және табиғи құбылыстардың өзгерістеріне бейім екенін дәлелдейді.

Бүгінгі таңда ЖИ технологияларының адам өмірінің барлық саласына, соның ішінде лингвистикалық кеңістікке де қарқынды енуі тіл ғылымының көкжиегін одан әрі кеңейтіп, сонымен қатар алдына жаңа міндеттер қойып жатыр. Үлкен тілдік модельдер (LLM) грамматикалық деңгейде сөйлем құрауды, логикалық операцияларды орындауды меңгергенімен, олардың қазақ тілінің ішкі иірімдерін, стилистикалық реңктері мен прагматикалық аспектілерін толық игеру деңгейі әлі де үлкен пікірталас тудыруда. Алгоритмдер миллиардтаған деректерді сұрыптай алады, алайда сан ғасырлар бойы мызғымас қалыпқа ие болған қазақ сөзінің астарындағы «жанды» мағынаны, ұлттық мінез бен контекстік ерекшеліктерді қаншалықты дәл қабылдайды? Осы сауал төңірегінде зерттеу жұмысымыздың негізгі бағыты айқындалады.

Зерттеудің өзектілігі – қазақ тілінің цифрлық әлеуетін бағамдау және ең жиі қолданылатын ЖИ жүйелерінің (ChatGPT, Gemini, Grok, DeepSeek) бірегей тілдік бірліктерді талдау қабілетін салыстырмалы түрде саралау. Мақаланың мақсаты – қазақ тілінің семантикалық, прагматикалық және стилистикалық ерекшеліктерін технологиялық сүзгі

ISSN 2788-7979 (online)

арқылы өткізе отырып, ЖИ құралдарының өнімділігіне объективті баға беру және ұлттық тілдің цифрлық болашағын айқындау.

Зерттеудің міндеттері:

– ЖИ модельдерінің қазақ тіліндегі лингвистикалық құзыреттілігін айқындап, артықшылықтары мен кемшіліктерін ажырату;

– ЖИ-ге этнонимдер мен ұлттық таным бірліктерін имплицитті түрде қолдана отырып мәтін құрастыру тапсырмаларын (промпттар) беру арқылы эксперимент жүргізу;

– ЖИ генерациялаған мәтіндерді табиғи тілдік деректермен салыстырып, технологиялық нәтижелер мен ғалымдардың іргелі тұжырымдарын бетпе-бет келтіре отырып, объективті ғылыми қорытынды жасау;

– Алынған нәтижелерге сандық және сипаттамалық көрсеткіштер арқылы баға беру.

Зерттеу нысаны ретінде ЖИ жүйелерінің қазақ тіліндегі мәтінді генерациялау үдерісі қарастырылады. Зерттеу пәні: Үлкен тілдік модельдердің ұлттық-мәдени семантиканы тану және лингвотанымдық тұрғыдан талдау әлеуеті. Зерттеу гипотезасы: ЖИ қазақ тілінің грамматикалық ерекшеліктері мен кең қолданыстағы тілдік бірліктерін сәтті игергенімен, мағыналық астары терең және халық арасында бәсеңдеген, көмескіленген сөздерін толық меңгере алмауы. Теориялық және практикалық маңызы: Ұлттық тілдің цифрлық кеңістіктегі даму жолын айқындай отыра, алынған нәтижелер ЖИ-ді қазақ тіліне бейімдеу және оқыту үдерісінде білім алушылардың тілдік сауаттылығы мен танымдық қабілеттерін дамыту бағытында таптырмас әдістемелік негіз болуын қамтамасыз ету.

ЖИ жүйелерінің қолданыстағы артықшылықтары мен кемшіліктері жөнінде шетелдік ғалымдар да көптеген елеулі ғылыми еңбектерін жарыса жариялауда. Әлемдік тілдердің қолданылу аясы мен цифрлық әлеуеті жоғары деңгейде болғандықтан, ЖИ қазақ тілінен гөрі оларды ұтымдырақ игере алатыны жасырын емес. Дегенмен ғалымдардың басым бөлігі «жасандылық» феномені тілдің тұрақталған, кодталған, берік келбетін қанша қайталап, өз ыңғайына бағындыруға тырысқанын алға тартса да, айтарлықтай өзгеріс алып келетініне күмәнмен қарайды.

Генеративті лингвистиканың негізін салушы Н. Хомский үлкен тілдік модельдердің (LLM) жұмыс жасау қағидасына сыни баға береді. Ғалым өзінің зерттеулерінде заманауи ЖИ алгоритмдерінің тек қана статистикалық ықтималдыққа сүйенетінін, сол арқылы тілдің үстіртін формаларына ғана шебер еліктейтінін дәйектейді. Оның пікірінше, бұл жүйелер адам танымына тән тілдің «терең құрылымына» және логикалық негізіне жете алмайды, сәйкесінше мәтінді грамматикалық тұрғыдан сауатты құрастырғанымен, оның шынайы табиғатын ұғынбайды [5].

Хомский атап көрсеткен технологиялық дәрменсіздік прагматикалық деңгейде қандай кедергілерге әкелетінін Й. Чойдың еңбектерінен көруге болады. Зерттеуші ЖИ-дің астарлы мағыналарды, метафоралық қолданыстарды және күнделікті «сауатты пайымдауды» (commonsense reasoning) қабылдауда айтарлықтай шектеулері бар екенін дәлелдеген [6, 143]. Адам пайымы белгілі бір сөйлем ішінде сөздердің жетіспеуі не айтылмай қалуы жағдаяттарына өзінің өмірлік тәжірибесіне сүйену арқылы шешімін таба алады. Алгоритмдер сансыз мәтінді өңдегенімен, адамға тән интуиция мен контекстік сезімталдықтан жұрдай болғандықтан, жасырын мағыналарды тікелей, тура мағынасында қабылдап, прагматикалық қателіктерге бой алдырады.

Аталмыш мәселелер көптілді ЖИ модельдерінде, әсіресе ағылшын тілінен басқа тілдерді өңдеу барысында тіпті күрделене түседі. Бір тілден екінші тілге аударма жасау – ЖИ-дің тілдерді меңгеру құзыреттілігін айқындауға болатын таптырмас амал. Бұл үрдісті технологиялық тұрғыдан С. Рудер бастаған ғалымдардың кросс-лингвистикалық зерттеулері айқындай түсті. Олар көптілді модельдердің басқа тілдерді жеке дара, тәуелсіз жүйе ретінде емес, басым мәдениеттің, көбінесе ағылшын тілінің, семантикалық үлгісіне салып талдайтынын анықтаған [7, 586]. Яғни, ЖИ басқа тілдегі мәтінді генерациялау барысында оның төл табиғатын емес, алгоритмнің базалық тіліндегі калькасын ұсынады. Мұндай

ISSN 2788-7979 (online)

технологиялық біріздендіру, яғни унификация ұлттық тілдердің лексика-семантикалық бірегейлігіне тікелей қауіп төндіреді.

Үлкен тілдік модельдердің қазақ тілі сияқты жалғамалы тілдердің ішкі контексін толық бере алмауының технологиялық себептері Р. Ахарони, М. Джонсон, О. Фират зерттеулерінде айқын көрініс тапқан. Ғалымдар ЖИ жүйелеріне жан-жақты баға беру мақсатында 100-ден астам тілге эксперимент жүргізіп, әр жұбына (қазақ-ағылшын, қазақ-испан) тиісті жеке модельдер құрастырған. Олардың мазмұндық қыры солғын болмасын деп, 1 миллионнан аса тілдік жағдаяттардың сүзгісінен өткізген. Бұл сынақтың нәтижесі аса өзекті ғылыми-техногендік жаңалық ашты. Егер модель қазақша →ағылшынша және ағылшынша → жапонша үйренсе, ол қазақша → жапонша аударуды да белгілі деңгейде орындай алады. Дегенмен зерттеушілердің тұжырымы бойынша, массалық көптілді модельдерде «сыйымдылық шектеу» орын алады. Соның салдарынан алгоритм аз ресурсты тілдердің күрделі морфологиялық құрылымы мен семантикалық астарын терең талдаудың орнына, оларды жоғары ресурсты тілдердің жеңілдетілген қалыбына бейімдеп жібереді [8, 3879].

Т. Наус тілдік модельдердің прагматикалық деңгейде «мәдени соқырлыққа» бой алдыруын ағылшын-араб тілдерін салыстыру арқылы көз жеткізген. Ғалым ЖИ-дің араб тіліндегі мәтіндерді өңдеу барысында ұлттық танымды қалай бұрмалайтынын «намаз және сусын» мысалы арқылы көрсеткен. Мәселен, алгоритмге исламдық намаздан кейінгі «going for a drink» (сусын ішуге бару) мәнмәтіні берілгенде, ол араб мәдениетіне тән «кофе немесе шайды» емес, батыстық үлгіге сүйеніп «құрамында спирті бар ішімдіктер» ұшырасатын мәтіндерді құрастырған. Бұл қателіктің түптөркіні – батыста «drink» сөзі статистикалық басым жағдайда «іс-шарадан соң көңіл көтеруге бару» секілді стандарттармен шендесуі. Сондықтан ғалым бұл эксперименттердің көрсеткіші ретінде ЖИ батыс дүниетанымына тәуелдігін және кейінгі уақытта ғана әлемдік діни-мәдени өзгешіліктерге бет бұратынын болжайды. Сонымен қатар араб тіліндегі адамның есімдері «кедейлік» және «қарапайымдылық» сөздерімен, ағылшын есімдері «байлық» ұғымымен жиі ассоциациялануын ЖИ-дің өзінде ұлттарға тағылған қисынсыз таптаурындардың бар екендігін айқындайды [9, 1].

МАТЕРИАЛДАР МЕН ӘДІСТЕР

Зерттеу материалы ретінде қазіргі соңғы үлгідегі жетекші жасанды интеллект модельдерінің (ChatGPT-5.3, Gemini 3, Grok 4 және DeepSeek-V3) генерациялаған мәтіндері қолданылады. Эмпирикалық базаны қалыптастыруда қазақ тілінің ұлттық танымын бейнелейтін, мәдени-әдеби қуаты бай әрі күнделікті қолданыста сирек ұшырасатын лексикалық бірліктерге басымдық беріледі. Атап айтқанда, ЖИ модельдерінің тілдік құзыреттілігін тексеру үшін күрделі фразеологизмдер, бейнелі теңеулер және ауыз әдебиетінің үлгілері негізге алынады.

Құрастырылған мәтіндердің басты ерекшелігі – модельдерге жіберілген арнайы промпттардың мақсатты түрде өңделген, тікелей нұсқаулары бар сұраныстар емес, имплицитті (жасырын) сауалдар арқылы берілуі. Бұл тәсіл ЖИ-дің жай ғана ақпаратты іздеп табу емес, тілдің жасырын мағыналық қатпарларын түсіну деңгейін айқындауға бағытталған.

Модельдердің лингвистикалық және прагматикалық әлеуетін бағалау үшін мынадай әдістер кешені қолданылады: ЖИ-дің берілген жағдаятқа бейімделуін тексеру үшін контекстік талдау; мәтіннің мағыналық тереңдігі мен адам танымына сәйкестігін анықтау үшін когнитивті-семантикалық талдау; сондай-ақ, жасырын интенцияны (түпкі ойды) дұрыс тану деңгейін өлшеу мақсатында прагматикалық талдау әдістері жүргізілді.

НӘТИЖЕЛЕР

Тіл білімінде фразеологизмдер – халықтың ғасырлар бойғы дүниетанымы, салт-дәстүрі мен мәдени коды сақталған ерекше құрылымдар. Әр халықтың өзіне тән фразеологиялық

ISSN 2788-7979 (online)

жүйесі оның ұлттық болмысын айқындайтын басты көрсеткіштердің бірі. Қазақ тіліндегі тұрақты тіркестер дистрибутивті-семантикалық қағида негізінде дамып, қоршаған ортадағы элементтердің логикалық өзара әрекеттесуі нәтижесінде жаңа концептуалды мағынаға ие болып, ажырамас құрылымдарға айналған. Мәселен, көшпелі мәдениетке тән ат әбзелдері, тұрмыстық бұйымдардың атаулары, табиғат құбылыстары уақыт өте келе адамның мінезін, психологиялық күйін немесе әлеуметтік қарым-қатынасын сипаттайтын күрделі семантикалық бірліктерге айналды. Ал бұл ЖИ-дің ең әлсіз тұсы. Сондықтан, эксперименттік кезеңнің басым бөлігі тұрақты тіркестерге арналды.

Қазақ тілінде «**Он төртінші туған айдай**» тұрақты тіркесі қолданылу жиілігі жағынан актив болса да, көпшілік этимологиялық-семантикалық мағынасын біле бермейді. «Сұлу, келбетті қыз» мағыналарына ие бұл тіркес не себепті айдың он төрт санымен байланысты? Бұл теңеуге қатысты жазушы М. Есламғали өз еңбегінде мынадай анықтама береді: «Наурыз күнін Батыс Қазақстанда 14 жұлдызда (наурыздың) атап өтеді. Мұның да өз түсінігі бар екен. Біріншіден, қазақ сұлу қыздарды суреттегенде, сипаттағанда «Он төртінші айдай боп» деп суреттейді. Шынында Ай әр айдың он төртінші жұлдызында шартабағай болып, әбден толысады. Сол себепті де халқымыз осы күнді «он төртінші жұлдыз» демейді, «айдың бір жаңасы» дейді. Яғни ықылым заманнан бері жаңа туған айдың алғашқы күні деп есептеп келгенді» [10, 69].

Тапсырма 1. «Он төртінші туған айдай» тіркесін қатыстырып 1 сөйлем құрастыр.

ChatGPT: *Жаңа түскен келін он төртінші туған айдай болып, ауылды сүйсіндірді.*

Gemini: *Оның жүзіндегі биязылық пен пәк сұлулықты көрген жұрт: «Бұл қыз нағыз он төртінші туған айдай екен!» — деп тамсанды.*

Grok: *Оның порымы он төртінші туған айдай көркем еді.*

DeepSeek: *Қыздың сұлулығы сонша, жүзі он төртінші туған айдай жарқырап, баршаның көңілін өзіне аударды.*

Сұраныс жауаптары көрсеткендей барлық ЖИ жүйелері бұл фразеологизмнің негізгі семантикасын дұрыс меңгергенін көрсетті. Атап айтқанда, барлық модельдердің жауаптарында тіркестің «толысқан айдай нұрлы, ерекше сұлу келбет» мағынасы сақталған. Сонымен қатар, олар тіркесті ұлттық танымға сай әйел бейнесін сипаттауда орынды қолданған. Жекелеген айырмашылықтар тек экспрессивтік реңк пен көркемдік деңгейінде байқалады: кейбір модельдерде (Grok) сипаттама ықшам берілсе, басқаларында (ChatGPT, Gemini, DeepSeek) бейнелілік пен эмоциялық бояу айқынырақ көрінеді.

«Айылын жимау» тіркесі «қорықпау», «қарсы әрекет ету» мағынасында жұмсалады. Академик Р. Сыздық бұл тіркестің астарын зерттей отырып: «Қазақ тілінде тек болымсыз мәнде айтылатын «айылын жимады» деген тіркес бар. Сырттан қарағанда мұндағы айыл сөзі тұрман әбзелінің бірінің аты сияқты. Ал шындығында айыл – моңғол тілінде «қорқыныш, үрей» деген мәндегі сөз. Сірә, бұл өте көне тіркес болу керек», – деген тұжырымды негіздейді [11, 26]. *Кейде бір тұрмыстың кем-кетігін еске алып, жоқтан өзгеге тарылып, шақ-шұқ көтеретін әйеліне шәу деп айылын жиятын Қозыбай жоқ (Қ. Жұмаділов).*

Тапсырма 2. «Айылын жимау» тіркесіне 2 сөйлеммен түсініктеме бер.

ChatGPT: *«Айылын жимау» — біреуге еркіндік бермей, үнемі қыспақта ұстау дегенді білдіреді. Сондай-ақ бұл тіркес аянып қалмай, тоқтамай әрекет ету мағынасында да қолданылады.*

Gemini: *«Айылын жимау» – сескенбеу, қаймықпау немесе ешкімді көзге ілмей, еркінсіп кету деген мағынаны білдіреді. Бұл тіркестің төркіні аттың айылын тартпай, бос жібергендегі еркіндігі мен асаулығына негізделген.*

Grok: *«Айылын жимау» — қазақтың тұрақты тіркесі, ешкімнен қорықпау, жасқанбау, сескенбеу деген мағына береді. Ол адамның еркін, батыл әрі қарсы әрекет етуге дайын тұрғанын білдіреді.*

DeepSeek: *«Айылын жимау» — айылды шешіп, аттан түсіп демалмау, яғни тоқтаусыз жүре беру деген тікелей мағынадан шыққан. Ауыспалы мағынада бұл тіркес адамның тынымсыз еңбек етіп, тынығуға уақыт таппай, бір істі үзбей орындауын білдіреді.*

ISSN 2788-7979 (online)

Тапсырма нәтижелері бойынша бұл жолы ЖИ модельдерінің жауаптарында мағыналық бірізділік сақталмай, әртүрлі интерпретациялар ұсынылғаны байқалады.

Gemini мен Grok модельдері «сескенбеу», «қорықпау», «жасқанбау» сияқты мағыналарды көрсету арқылы фразеологизмнің негізгі мәніне біршама жақындайды. Алайда, олардың түсіндірулерінде «қарсы әрекет ету» реңкі толық ашылмаған.

ChatGPT жауабында «қыспақта ұстау» мағынасы бұл тіркеске мүлдем жат. Бұл жерде модель «айылын жию» (сескену) мен «айылын жиғызу» (бетін қайтару) ұғымдарын шатастырып алғандай. Ал DeepSeek «айыл» сөзін тек аттың әбзелі ретінде қабылдап, тіркесті «демалмау, еңбек ету» деп өз бетінше пайымдаған. Бұл контекстке де, сөздік қорға да сай келмейді.

Жаққұрт келсін – қазақта қарғыс, реніш мәнді қолданылатын тұрақты тіркес. «Құдай жағыңнан алсын, зар жақ боп сөйлей алмай қал! (Тісіңнің құрты ұстасын деген мағынадан келіп шыққан)» [12, 176]. Халық арасында емі жоқ жаман аурудың бірі «жаққұртты» шақыру біреуге қатты өкпелі, қапалы болған кезде қолданылған екен.

Тапсырма 3. «Жаққұрт келсін» тіркесімен 1 сөйлем құрастыр.

ChatGPT: *Ашуланған ол: «Жаққұрт келсін!» — деп, теріс бұрылып кете барды.*

Gemini: *Қарт ақсақал жас жігіттің ел бірлігі туралы айтқан парасатты сөзін естігенде: «Бәрекелді, балам, айтқаныңа жаққұрт келсін, мәртебең биік болсын!» — деп батасын берді.*

Grok: *Жаңа үйге кіргенде «Жаққұрт келсін!» деп тіледі.*

DeepSeek: *Жаныңа жаққұрт келсін, неге менің айтқанымды орындамайсың!*

Талдау нәтижесі бойынша ұсынылған жауаптардың прагматикалық реңктерінде елеулі айырмашылықтар байқалды.

Атап айтқанда, ChatGPT және DeepSeek модельдері тіркесті жағдаятқа сәйкес, қарғыс мәнінде орынды қолданған. Бұл олардың фразеологизмнің эмоциялық-экспрессивтік сипатын ішінара дұрыс меңгергенін көрсетеді.

Ал Gemini мен Grok модельдерінің жауаптарында семантикалық қателік байқалады: тіркес жағымды тілек, бата мәнінде қолданылып, оның бастапқы қарғыс мағынасы бұрмаланған. Бұл жағдай ЖИ жүйелерінің ұлттық-мәдени кодқа тән прагматикалық реңктерді ажыратуда қиындықтарға тап болатынын айқын көрсетеді.

Қазақ халқының салт-дәстүрлері – ұлттық болмыстың өзегі мен ғасырлар бойы қалыптасқан рухани мұраның бірегей жиынтығы. Әрбір әдет-ғұрыптың астарында терең тәрбиелік мән мен халықтың даналығы жатқандықтан, олар ұлттық «феномен» ретінде қайталанбас бағалы қазынаға айналып, уақыт өтсе де маңызын жоғалтпайды. Соған сай қазіргі кезде ЖИ жүйелерінің қазақтың салты, әдет-ғұрпы, жөн-жорамалдарын білу мүмкіндігі цифрлық кеңістікте ақпараттың күннен-күнге артып келе жатқандығымен тығыз байланысты. Көпшілікке кеңінен таныс «тұсаукесер», «ашамайға мінгізу», «сырға салу», «байғазы» сияқты дәстүрлерді ЖИ жақсы меңгерген. Ал кей салты қарбалас заманның өткіншілігі салдарынан тіпті орта жастағы ел жұртының есінен ұмыт болуына шақ. Бұған халқымыздың **«Көгентүп»** деген ерекше салтын жатқызсақ болады. Атақты қоғам қайраткері, жазушы З. Ахметова «Күретамыр» еңбегінде: «Көгентүп дегеніміз – балаға берілетін қозы-лақ, бота, бұзау немесе құлын. Бүгінгілер біле бермейді, «көген» – қозыларды байлайтын ұшы түймеленіп түйілген қысқа жіптер. Оларды жуан арқанға тағып, қозыларды көгендеген» деген түсінік берген [13 93]. Шынында, қазір қоғам бұрынғыдай төрт түлікті сыйлық ретінде табыстамайды. Содан болуы керек, бұл көрнекті дәстүрге тіпті «көген» тәрізді жіптің қатысы бар екенін де ұмытылуы уәжді құбылыс.

Тапсырма 4. «Көгентүп» деген сөзді қатыстырып 1 сөйлем құрастыр.

ChatGPT: *Кешке қарай қозыларды көгентүпке байлап қойды.*

Gemini: *Нағашы атасы жиеніне алғаш ат мінгенде: «Мынау ақбас қозы сенің көгентүбің болсын», — деп батасын берді.*

Grok: *Туысқанымыздың үйіне барғанда, олардың кішкентай баласына көгентүп ретінде бір қозы атап бердік.*

ISSN 2788-7979 (online)

DeepSeek: Ғарышкер ашық ғарышқа шығарда көгентүпті киді.

Алынған жауаптарда модельдердің бұл ұғымды игеру деңгейі әркелкі екені байқалады. Дәлірек айтқанда, Gemini және Grok модельдері «көгентүп» сөзін ұлттық-мәдени контексте сәтті қолданып, оның «сыйлыққа берілетін мал» мағынасын сақтай алған. Бұл олардың этнографиялық мазмұнды ішінара дұрыс меңгергенін көрсетеді.

Ал ChatGPT жауабында «көгентүп» сөзі «қозыларды байлайтын құрал» ретінде қате интерпретацияланған. Бұл жағдайда «көген» мен «көгентүп» ұғымдарының ара-жігі ажыратылмай, семантикалық шатасу орын алған.

DeepSeek моделінің жауабында ең өрескел қателікке жол берілген, «көгентүп» сөзін мүлдем басқа салаға жатқызылып, оны ғарышкердің киімі (скафандр) ретінде сипатталған. Бұл жүйенің базасында қазақ тілінің сирек кездесетін этнографиялық сөздік қорының өте төмен деңгейде екенін байқатады.

Қазақтың шешендік өнері – қысқа да нұсқа, бейнелі, ырғақты әрі импровизацияға негізделген ауыз әдебиетінің ең мәнді жанрларының бірі. Бұл салада бір ауыз сөзбен дау шешіліп, терең философиялық ой метафора арқылы беріледі. Ол халықтық даналық пен мәдени кодтардың шоғырланған көрінісі іспетті.

Шешендік өнердің күрделілігі тілдің көпқабаттылығымен, сондай-ақ көшпелі дүниетаным мен контекстік тәуелділікпен айқындалады. Би-шешендердің сын сағаттағы жауабы тілдік импульске негізделіп, нақты жағдаятқа байланысты туындайды. Осыған орай, ЖИ модельдері дәл осы жанрды игеруде едәуір қиындықтарға тап болуы мүмкін, өйткені ауыз әдебиеті толық цифрланбаған. Сонымен қатар, тілдік мұра болғандықтан оның табиғи құрылымын өзгертуге немесе еркін түрлендіруге болмайды.

Танымал Сырым батыр мен Хиуа ханының арасындағы: *«Ат бәйгесі мен палуан күресінен жеңілген хан, ең болмаса, сөзден сүріндірейін деп, Сырымға үш сұрақ береді:*

– Жігіт-мігіт деген не сөз?

Ат-мат деген не сөз?

Ас-мас деген не сөз?

– Жыққан – жігіт, жығылған – мігіт,

Озған – ат, қалған – мат.

Жал-жая – ас, шөп-шөлең – мас, – дейді Сырым», – бұл оқиға ауыз әдебиетінің жарқын үлгісі ретінде бағаланады [14 193]. Қойылған сұрақтардың астарлылығына қарамастан, батыр сөз сайысында мүдірмей, жауапты қысқа әрі сенімді түрде қайырады.

Тапсырма 5. Сырым батырдың жауабы қандай болды? 3 сөйлеммен түсініктеме бер. (Тапсырма ханның сұрағымен шектелді).

ChatGPT: *Сырым Датұлы бұл сұрақтарға төмендегідей жауап береді: Жігіт – ел қорғайтын нағыз ер, ал мігіт – соған ұқсас, бірақ шын жігіт емес; ат – ер қанаты, ал мат – жай көлік. Ас – адал тағам, ал мас – қадірсіз, орынсыз ас.*

Gemini*: Сырым шешен ханның сұрағына: «Еліне қорған болған азамат — Жігіт, оның соңынан ерген көлеңкесі — Мігіт», — деп жауап береді. Ал алысқа шапқан тұлпарды — Ат, оған ілесе алмайтын мәстекті — Мат деп түсіндіреді. Сонымен қатар, қадірлі дәмді — Ас, ал оның жұғын-сарқытын — Мас деп, құндылық пен сапасыздықтың аражігін ажыратып береді.*

Grok: *Сырым ханның үш сұрағына былай жауап берді:*

Жыққан — жігіт, жығылған — мігіт.

Озған — ат, қалған — мат.

Жал-жая — ас, шөп-шөлең — мас.

Осы үш қысқа жауаппен Сырым ханды сөзден жеңіп шықты.

DeepSeek: *Мұнда Сырым Датұлы ханның үш сұрағына былай жауап береді: «Жігіт – азамат, мігіт – жалғыз ат», «Ат – тұлпар, мат – бәйгеде озып кеткен», «Ас – тамақ, мас – тойған адам». Осылайша ол әр сөздің қосымшасына да жеке мағына беріп, ханның сөзден сүріндіру әрекетін тапқырлықпен жеңеді. Бұл жауап арқылы Сырым өзінің шешендік өнері мен ақылдылығын дәлелдейді.*

ISSN 2788-7979 (online)

Тапсырма нәтижесіне сай, ЖИ модельдерінің жауаптары күтілгендей алуан сипатта болды. Grok моделі түпнұсқа мәтінге ең жақын және дәл жауап беріп, шешендік сөздің ырғақты құрылымын сақтаған. Бұған ақпараттық кеңістікте іздеу қабілеті жақсы септігін тигізгенін жатқызсақ болады.

Ал ChatGPT мен Gemini-дің «интерпретациялау» әрекеті керісінше, фольклорлық мәтіннің түпнұсқалығын бұзды. Олар сұрақтың астарын «түсіндіруге» тырысып, қысқа жауапты ұзақ, түсініктеме сипатындағы мәтінге айналдырды. Бұл – классикалық галлюцинацияның көрінісі. Модельдер оқырман жауапты тексермеуі мүмкін екенін ескеріп, өзінің жалпылама біліміне сүйеніп, стильдік жағынан ұқсас ойдан шығарылған мәтін құрастырған. Нәтижесінде лексикалық форма мен ырғақ жоғалған. Алайда, бұл екі модельдің мақтайтын тұсы жауаптары логикалық тұрғыдан ақиқатқа жақындау. Семантикалық дәлдік ескеріліп, яғни антонимдік жұптар ұтымды (*Ас – адал тамақ, мас – орынсыз ас; ат – тұлпар, мат – мәстек*) уәжделе алған.

DeepSeek мәтінді мүлдем қате түсініп, «мат – бәйгеде озып кеткен», «мас – тойған адам» деген сияқты логикалық және семантикалық ауытқуларға жол берген. Яғни, ЖИ-ге тән мәтін мазмұнын сөзбе-сөз аша салуға тырысқан.

ТАЛҚЫЛАУ

Зерттеу нәтижелерін объективті тұрғыдан бағалау мақсатында кестеге төмендегідей критерий бойынша сандық көрсеткіштерді енгіздік:

3 балл – толық дұрыс (семантика, прагматика, мәдени код сақталған);

2 балл – жартылай дұрыс (негізгі мағына дұрыс, бірақ реңктік мәні жетімсіз);

1 балл – қате интерпретация (кейбір элементтер дұрыс, бірақ жалпы ауытқу бар);

0 балл – толық қате / галлюцинация.

1-кесте. ЖИ модельдерінің ұлттық-мәдени лексиканың меңгеру көрсеткіштері

Тапсырма	Мысалдар	ChatGPT	Gemini	Grok	DeepSeek	Дұрыс жауаптар саны
1	«Он төртінші туған айдай»	3	3	3	3	4/4
2	«Айылын жимау»	0	2	2	0	2/4
3	«Жаққұрт келсін»	3	0	0	3	2/4
4	«Көгентүп»	1	3	3	0	2/4
5	Сырым батырдың жауабы	1	1	3	0	1/4
Жалпы балл	—	8	9	11	6	34/60 (57 %)
Орташа балл	—	1,6	1,8	2,2	1,2	—
Дәлдік %	—	53 %	60 %	73 %	40 %	—

Қазақ тіліндегі этномәдени лексиканың ЖИ модельдерімен игерілу деңгейін талдау нәтижесінде жалпы дәлдік 57% құрады. Бұл көрсеткіш ЖИ жүйелерінің қазақтың ұлттық-мәдени кодын, әсіресе көшпелі дүниетанымға тән тұрақты тіркестер мен салт-дәстүрлерді толық меңгеруде әлі де шектеулі мүмкіндіктері бар екенін айқын көрсетеді.

Салыстырмалы талдау барысында ең жоғары нәтижені Grok көрсетті (73 %). Аталған модель фразеологизмдердің семантикалық өзегін сақтаумен қатар, фольклорлық мәтіндердің ішкі ырғақтық және экспрессивтік табиғатын (әсіресе Сырым батыр жауабына қатысты тапсырмада) неғұрлым дәл жеткізе алды. Бұл оның контекстік интерпретациялау мүмкіндігінің жоғары деңгейде екенін және қазақ тіліне қатысты дерекқорының біршама дамығанын аңғартады.

Келесі сатыны Gemini 60% көрсеткішпен иеленді. Ол «он төртінші туған айдай», «көгентүп» тәрізді этнографиялық мазмұндағы бірліктерді дұрыс интерпретациялағанымен,

ISSN 2788-7979 (online)

прагматикалық реңктерді, әсіресе қарғыс немесе экспрессивтік мәндерді айқындауда дәлсіздікке жол берген.

ChatGPT 53% деңгейінде орташа нәтиже көрсетті. Бұл модель жекелеген фразеологизмдердің эмоциялық-экспрессивтік сипатын («жаққұрт келсін») салыстырмалы түрде дұрыс жеткізгенімен, кейбір тіркестердің («айылын жимау») мағынасын бұрмалап, контекстік сәйкессіздікке ұрынған.

Ең төменгі көрсеткіш DeepSeek-ке тиесілі болды (40 %). Бұл жүйе сирек қолданылатын этнографиялық лексиканы («көгентүп») және фольклорлық мазмұндағы мәтіндерді интерпретациялау барысында өрескел қателіктерге жол беріп, галлюцинациялық сипаттағы жауаптар ұсынған.

Жалпы алғанда, қазақ фразеологиясы мен дәстүрлі мәдениетке қатысты тілдік бірліктерді ЖИ көмегімен талдау барысында модельдердің семантикалық дәлдігі белгілі бір деңгейде қалыптасқанымен, прагматикалық және мәдени-этнографиялық қабаттарды меңгеру деңгейі жеткіліксіз екені анықталды. Бұл жағдай қазақ тілінің ұлттық корпусын, фразеологиялық қорын және ауыз әдебиеті үлгілерін жүйелі түрде цифрландыру мен оларды ЖИ модельдерін оқытуға кеңінен енгізудің өзектілігін көрсетеді. Grok моделінің салыстырмалы басымдығына қарамастан, жалпы алғанда, қазіргі ЖИ жүйелері қазақ халқының ғасырлар бойы қалыптасқан дүниетанымдық моделін толық әрі дәл бейнелеу үшін әлі де жетілдіруді қажет етеді.

ҚОРЫТЫНДЫ

Бұл зерттеу жұмысында қойылған негізгі мақсат пен міндеттерге толық қол жеткізілді: заманауи жасанды интеллект модельдерінің (ChatGPT, Gemini, Grok, DeepSeek) қазақ тілінің семантикалық, прагматикалық және этнолингвистикалық бірліктерін өңдеу әлеуеті кешенді түрде сараланды. Эксперимент барысында ғалымдардың ЖИ технологияларына қатысты сыни көзқарастары толығымен өз дәлелін тапты. Табиғи тілдік үлгі мен нақты дәлел ретінде қазақ ғалымдары мен жазушыларының тұжырымдары алынып, ЖИ модельдерінің жауаптарымен салыстырылды. Зерттеушілердің «ЖИ тілдің терең құрылымы мен логикалық негізін ұғынбайды» деген пікірлері және «мәдени соқырлық» концептісі қазақ тіліне қатысты айқын көрініс берді. Жүйелер грамматикалық тұрғыдан сауатты сөйлемдер құрастырғанымен, сирек қолданылатын тіркестердің бойындағы лингвистикалық ерекшеліктерді ажырата алмай, оларды батыстық стандарттарға немесе тікелей аудармалық калькаға салып бұрмалады.

Жүргізілген тәжірибе ЖИ-дің қазақ тілін әлі де дербес, толыққанды когнитивті жүйе ретінде қабылдамайтынын көрсетті. Модельдердің жалпы дәлдігі 57%-ды құрауы келесідей маңызды тұжырымдар жасауға негіз болады: ЖИ қазақ тілінің жалпыхалықтық белсенді лексикасын біршама жақсы меңгерген, алайда ауыз әдебиетінің үлгілері, тұрақты тіркестер мен этнографиялық атаулар келгенде алгоритмдік галлюцинацияға ұшырайды. Яғни, ЖИ модельдері қазақ сөзінің «жанын» – оның прагмастилистикалық реңкі мен ұлттық-мәдени кодын тануда дәрменсіздік танытып отыр.

Жасанды интеллектінің болашақ даму алғышарттары тек техникалық параметрлерді (миллиардтаған таңбаларды) арттырумен шектелмеуі тиіс. Басты көзделген мақсат – ЖИ-ді басқа әлемде танымал тілдерден аударылған дайын пішіні емес, қазақ халқының төл тарихы, мәдениеті және әдебиеті негізінде жасалған таза ұлттық корпустармен оқыту.

Цифрлық дәуірде филолог ғалымдардың миссиясы енді тек ЖИ-дің әлсіз тұстарын сынаумен шектелмеу керек. Ғылыми зерттеу әдістерінің талаптарына сүйене отырып, біз тілдің цифрлық кеңістіктегі прагмастилистикалық сипатын жақсартуға нақты қадамдар жасауымыз қажет. Зерттеуші ғалымдардың ендігі қадамдары ЖИ-ді оқытуға жарамды таза, сапалы әрі ұлттық танымға бай мәтіндік базаларды іріктеп, алгоритмдерге «адамдық пайымды» үйретуге бағытталуы шарт. Осы мақсатта жоғары академиялық жетекшілікпен ЖИ тілінің прагмастилистикасын зерттейтін арнайы цифрлық платформалар, танымдық сайттар

ISSN 2788-7979 (online)

мен бейнеарналар құрып, тілдік кодтың сандық кеңістікте дұрыс таңбалануын қамтамасыз ету – кезек күттірмейтін басты назардағы мәселе.

ӘДЕБИЕТТЕР ТІЗІМІ

1 Цифрландыру және жасанды интеллект жылын жариялау туралы [Электрондық ресурс] // Қазақстан Республикасы Президентінің ресми сайты. – Астана, 2026. – 6 қаңтар. – URL: https://www.akorda.kz/kz/cifrlandyru-zhane-zhasandy-intellekt-zhylyn-zhariyalau-turaly-601458 (қаралған күні: 14.03.2026).
2 Sapir E. The status of linguistics as a science // Language. – 1929. – Vol. 5, № 4. – P. 207–214.
3 Қайдаров А. Т. Қазақтар ана тілі әлемінде (этнолингвистикалық сөздік) Адам. – Алматы: Дайк-Пресс, 2009. – 784 б.
4 Сыздықова Р. Қазақ әдеби тілінің тарихы (XV–XIX ғасырлар). – Алматы : Ана тілі, 1993. – 319 б.
5 Chomsky N. Noam Chomsky: The false promise of ChatGPT [Электрондық ресурс] / N. Chomsky // The New York Times. – 2023. – URL: https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (қаралған күні: 7.03.2026).
6 Choi Y. The curious case of commonsense intelligence // Daedalus. – 2022. – Vol. 151, № 2. – P. 139–155. https://doi.org/10.1162/daed_a_01906.
7 Ruder S., Vulić I., Søgaard A. A survey of cross-lingual word embedding models // Journal of Artificial Intelligence Research. – 2019. – Vol. 65. – P. 569–631. https://doi.org/10.1613/jair.1.11640
8 Aharoni R., Johnson M., Firat O. Massively multilingual neural machine translation // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). – Minneapolis, 2019. – P. 3874–3884. https://doi.org/10.18653/v1/N19-1388
9 Naous T., Ryan M. J., Ritter A., Xu W. Having beer after prayer? Measuring cultural bias in large language models // Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024). – Bangkok, 2024. – Vol. 1. – P. 16366–16393. https://doi.org/10.18653/v1/2024.acl-long.862
10 Есламғалиұлы М. Сан қилы сауал. – Алматы: Ана тілі, 1991. – 86 б.
11 Сыздықова Р. Сөздер сөйлейді (сөздердің қолданылу тарихынан). Көптомдық шығармалар жинағы. – Алматы: Ел-шежіре, 2014. – 412 бет
12 Қазақ тілінің фразеологиялық сөздігі / құраст. І. К. Кеңесбаев. – Алматы : Ғылым, 1977. – 712 б.
13 Ахметова З. Күретамыр. – Алматы, 2018. – 208 бет.
14 Қазақтың шешендік өнері. Алматы: «Мазмұндама» қоғамдық қоры, 2022. – 228 бет.

Материал 29.03.2026 баспаға түсті

Искусственный интеллект и национально-культурная идентичность казахского языка: проблемы семантической и прагматической интерпретации

Е.А. Амангельдин[1], А.С. Мусаев[1], Д.А. Алкебаева[1]
[1]Казахский национальный университет имени аль-Фараби, Алматы, 050000, Казахстан

В условиях современного цифрового развития природа и функции языка переосмысливаются в новой технологической среде. В частности, стремительное развитие систем искусственного интеллекта (ИИ) сделало оценку цифрового потенциала национальных языков актуальной научной проблемой. Настоящее исследование направлено на определение степени освоения этнокультурной лексики казахского языка моделями ИИ. Основная цель состоит в анализе семантических, прагматических и стилистических особенностей казахского языка с помощью систем ИИ и сравнительной оценке способности моделей распознавать национально-культурный код. Ключевая идея исследования основана на

ISSN 2788-7979 (online)

понимании того, что язык – это не только структурная система, но и когнитивный феномен, отражающий национальное мировоззрение, что позволяет оценить компетентность ИИ в его обработке.

Научная и практическая значимость исследования заключается в определении перспектив развития казахского языка в цифровом пространстве и обосновании необходимости адаптации моделей ИИ к национальному языку. Методологическую основу составляют контекстный, когнитивно-семантический и прагматический методы анализа, реализованные посредством экспериментального подхода.

Результаты эксперимента показали, что модели ИИ в среднем демонстрируют 57% эффективности в распознавании национально-культурного кода казахского языка. Среди моделей наивысший результат показал Grok (73%), в то время как DeepSeek продемонстрировал самый низкий показатель (40%). Несмотря на достаточно высокий уровень грамматической корректности, модели допускают систематические ошибки при интерпретации прагматических оттенков и этнографических значений, а также проявляют случаи алгоритмических галлюцинаций.

Ценность исследования состоит в комплексном рассмотрении национально-культурных особенностей казахского языка в контексте цифровых технологий. Полученные результаты имеют практическое значение для адаптации систем ИИ к казахскому языку, расширения лингвистических корпусов и эффективного применения этих технологий в образовательном процессе.

Ключевые слова: искусственный интеллект, большие языковые модели, этнокультурная лексика, национальный код, когнитивная лингвистика, прагматика, семантика, стилистика.

СПИСОК ЛИТЕРАТУРЫ

1 Официальный сайт Президента Республики Казахстан. [Электрон. ресурс] // Об объявлении Года цифровизации и искусственного интеллекта. – Астана, 2026. – 6 января. URL: https://www.akorda.kz/ru/ob-obyavlenii-goda-cifrovizacii-i-iskusstvennogo-intellekta-601222 (дата обращения: 14.03.2026).

2 Sapir E. The status of linguistics as a science // Language. – 1929. – Vol. 5, № 4. – P. 207–214.

3 Кайдаров А.Т. Казахи в мире родного языка (этнолингвистический словарь) Человек. – Алматы: Дайк-Пресс, 2009. – 784 б.

4 Сыздыкова Р. История казахского литературного языка (XV–XIX века). – Алматы: Ана тілі, 1993. – 319 с.

5 Chomsky N. Noam Chomsky: The false promise of ChatGPT. [Электрон. ресурс] // The New York Times. – 2023. – URL: https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (дата обращения: 7.03.2026).

6 Choi Y. The curious case of commonsense intelligence // Daedalus. – 2022. – Vol. 151, № 2. – P. 139–155. https://doi.org/10.1162/daed_a_01906.

7 Ruder S., Vulić I., Søgaard A. A survey of cross-lingual word embedding models // Journal of Artificial Intelligence Research. – 2019. – Vol. 65. – P. 569–631. https://doi.org/10.1613/jair.1.11640

8 Aharoni R., Johnson M., Firat O. Massively multilingual neural machine translation // Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT). – Minneapolis, 2019. – P. 3874–3884. https://doi.org/10.18653/v1/N19-1388

9 Naous T., Ryan M. J., Ritter A., Xu W. Having beer after prayer? Measuring cultural bias in large language models // Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (ACL 2024). – Bangkok, 2024. – Vol. 1. – P. 16366–16393. https://doi.org/10.18653/v1/2024.acl-long.862

10 Есламгалиулы М. Разнообразные вопросы. – Алматы: Ана тілі, 1991. – 86 с.

11 Сыздыкова Р. Слова говорят (из истории употребления слов). Собрание многотомных произведений. – Алматы: Ел-шежіре, 2014. – 412 с.

ISSN 2788-7979 (online)
12 Фразеологический словарь казахского языка / сост. І. К. Кеңесбаев. – Алматы, Издательство «Ғылым» Казахской ССР, 1977. – 712 с.
13 Ахметова З. Күретамыр. – Алматы, 2018. – 208 с.
14 Ораторское искусство казахов. – Алматы: Общественный фонд «Мазмұндама», 2022. – 228 с.

Материал поступил в редакцию журнала 29.03.2026

Artificial Intelligence and the National-Cultural Identity of the Kazakh Language: Issues of Semantic and Pragmatic Interpretation

Y. Amangeldin[1], A. Mussayev[1], D. Alkebaeva[1]
[1]Al-Farabi Kazakh National University, Almaty, 050000, Kazakhstan

In the context of contemporary digital development, the nature and functions of language are being reconsidered within a new technological environment. In particular, the rapid advancement of artificial intelligence (AI) systems has made the evaluation of the digital potential of national languages a pressing scientific issue. This study aims to determine the extent to which AI models can master the ethnocultural vocabulary of the Kazakh language. The primary objective is to analyze the semantic, pragmatic, and stylistic features of the Kazakh language through AI systems and to comparatively assess their ability to recognize the national-cultural code. The core idea of the research is based on the understanding that language is not merely a structural system, but a cognitive phenomenon that reflects the national worldview, thereby enabling an assessment of AI's competence in processing it.

The scientific and practical significance of the study lies in identifying the prospects for the development of the Kazakh language in digital space and in substantiating the need to adapt AI models to the national language. Methodologically, the research relies on contextual, cognitive-semantic, and pragmatic analysis methods, implemented through an experimental approach.

The experimental results indicate that AI models demonstrate an average performance of 57% in capturing the national-cultural code of the Kazakh language. Among the models tested, Grok achieved the highest accuracy (73%), while DeepSeek showed the lowest performance (40%). Despite a sufficient level of grammatical correctness, the models exhibited systematic errors in interpreting pragmatic nuances and ethnographic meanings, along with instances of algorithmic hallucinations.

The value of the study lies in its comprehensive examination of the national-cultural features of the Kazakh language in the context of digital technologies. The findings have practical significance for adapting AI systems to the Kazakh language, expanding linguistic corpora, and effectively integrating these technologies into the educational process.

Key words: artificial intelligence, large language models, ethnocultural vocabulary, national code, cognitive linguistics, pragmatics, semantics, stylistics.

REFERENCES

1 President of the Republic of Kazakhstan (2026) *Sifrlandyru jäne jasandy intelekt jylyn jarialau turaly* [On declaring the Year of Digitalization and Artificial Intelligence]. Astana, 6 January. Available at: https://www.akorda.kz/kz/cifrlandyru-zhane-zhasandy-intellekt-zhylyn-zhariyalau-turaly-601458 (Accessed: 14 March 2026).
2 Sapir, E. (1929) "The status of linguistics as a science", *Language*, 5(4), pp. 207–214.
3 Kaidarov, A.T. (2013) Q*azaqtar ana tılı älemında: etnolingvistikalyq sözdık.* [Kazakhs in the world of their native language: Ethnolinguistic dictionary]. Almaty: Daik-Press.
4 Syzdykova, R. (1993) *Qazaq ädebi tılınıñ tarihy (XV–XIX ğasyrlar)* [History of the Kazakh Literary Language (15th–19th Centuries)]. Almaty: Ana Tili,

ISSN 2788-7979 (online)
5 Chomsky, N. (2023) “Noam Chomsky: The false promise of ChatGPT”, *The New York Times*. Available at: https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html (Accessed: 7 March 2026).
6 Choi, Y. (2022) “The curious case of commonsense intelligence”, *Daedalus*, 151(2), pp. 139–155. doi: https://doi.org/10.1162/daed_a_01906.
7 Ruder, S., Vulić, I. and Søgaard, A. (2019) “A survey of cross-lingual word embedding models”, *Journal of Artificial Intelligence Research*, 65, pp. 569–631. doi: https://doi.org/10.1613/jair.1.11640.
8 Aharoni, R., Johnson, M. and Firat, O. (2019) “Massively multilingual neural machine translation”, in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota. Association for Computational Linguistics, pp. 3874–3884. doi: https://doi.org/10.18653/v1/N19-1388
9 Naous, T., Ryan, M.J., Ritter, A. and Xu, W. (2024) “Having beer after prayer? Measuring cultural bias in large language models”, in *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Bangkok, Thailand. Association for Computational Linguistics, pp. 16366–16393. doi: https://doi.org/10.18653/v1/2024.acl-long.862
10 Eslamgaliuly, M. (1991) *San qily saual* [Various questions]. Almaty: Ana tili.
11 Syzdykova, R. (2014) *Sözder söileidı (sözderdıñ qoldanylu tarihynan). Köptomdyq şyğarmalar jinağy* [Words speak (from the history of word usage). Collected works in multiple volumes]. Almaty: El-shezhire.
12 Kenesbaev, I.K. (1977) *Qazaq tılınıñ frazeologialyq sözdıgı* [Kazakh phraseological dictionary]. Almaty: Publishing House of the Kazakh SSR.
13 Akhmetova, Z. (2018) *Küretamyr* [Main artery]. Almaty.
14 Qazaqtyñ şeşendık önerı [Kazakh oratory art] (2022). Almaty: «Mazmūndama» qoğamdyq qory

Received: 29.03.2026

ҒТАМР 17.00.00 DOI: 10.59102/kufil/2026/iss3pp256-267

А.М. Ахмеджанова [1], Г.К. Кажибаева[1]
[1]Ш. Уәлиханов атындағы Көкшетау университеті, Көкшетау, 020000, Қазақстан

ТӨЛЕГЕН ҚАЖЫБАЙ ПОЭЗИЯСЫНЫҢ КӨРКЕМДІК-ФИЛОСОФИЯЛЫҚ ЕРЕКШЕЛІКТЕРІ

Зерттеудің мақсаты – Төлеген Қажыбай поэзиясындағы өмір–өлім, уақыт–мәңгілік және туған жер–тарихи жад семантикалық модельдерінің көркемдік қызметін айқындап, ақынның ұлттық поэтикалық дәстүрдегі даралығын салыстырмалы-типологиялық тұрғыдан негіздеу. Материал ретінде Т. Қажыбайдың «Өмір жыры» циклі мен «Көкшеге» өлеңі, ал салыстырмалы контекст ретінде М. Мақатаевтың «Өмір дейтін», «Туған жерім» және Т. Айбергеновтің «Өмір», «Дала» өлеңдері алынды. Салыстыру тақырыптық өзек, лирикалық субъект, кеңістік бейнесі, жетекші троптар және композициялық-семантикалық құрылым өлшемдері бойынша жүргізілді. Нәтижесінде үш ақынға өмірдің қозғалыс, қарама-қайшылық және рухани сабақтастық ретінде бейнеленуі ортақ екені, ал Т. Қажыбайдың даралығы дәстүрлі толғау, арнау және жоқтау поэтикасын Көкше кеңістігімен, ұлттық тарихи жадпен және тұлғалық рефлексиямен ұштастыруынан көрінетіні анықталды. Ақын поэзиясында антитеза, метафора, символ, психологиялық параллелизм мен композициялық шеңбер жекелеген тәсілдер емес, адам–уақыт–мекен тұтастығын қалыптастыратын өзара сабақтас көркемдік жүйе ретінде қызмет етеді.