

С.Қ. Ыдырыс¹, Л.Ж. Бейсенбаева^{2*}, Д.Ж. Оспанова³

¹Шайсұлтан Шаяхметов атындағы «Тіл – Қазына» ұлттық ғылыми-практикалық орталығы, Астана, Қазақстан;
^{2,3} Л.Н. Гумилев атындағы Еуразия ұлттық университеті, Астана, Қазақстан
(e-mail: omarovask@mail.ru, tancopan@gmail.com, dana.zhanabekkyzy@gmail.com)

Корпус мәтіндеріндегі грамматикалық форма жиілігі (<https://qazcorporaz.kz> негізінде)

Мақалада қазақ тілінің грамматикалық аспектілерін зерттеу үшін корпус материалының мүмкіндіктерін қаншалықты қажетімізге жарата аламыз деген мәселе аясында етістіктердің грамматикалық формалары жиілігі қарастырылады. Қазақ тілінің сөз жиілігі мәселесі ғылыми объекті ретінде зерттеліп, сөздіктер құрастырылғанымен, осы мәселеде қазақ тілінің корпус базасы мүмкіндіктері толыққанды зерттеулер жүргізу үшін пайдаланылмай келе жатыр, өйткені қазақ тілінің ұлттық корпусын жасау мәселесі кеш қолға алынды. Корпус базасындағы мәтіндерді тілдік зерттеулер үшін пайдалану өте маңызды. Корпус мәтіндерін зерттеу тілдің жүйесіндегі лексикалық қабаттың, синтаксистік конструкциялардың, грамматикалық формалардың өзгерісін, дамуын анықтауға жол ашады. Кез келген тілдің сөздік қоры құбылмалы, өзгермелі болатындықтан, әрбір жылда тіл қандай сөздермен, сөзформаларымен толығады, өзге тілдерден енген сөздерге деген тілдің иммунитеті қандай, жаңа сөзформаларын қандай күйде қабылдады деген мәселеде корпус базасы жыл сайын толықтырылып, жетілдіріліп отырғаны абзал. Бұндай зерттеу жүргізілмесе, тіліміздің даму бағыты бұлыңғырланады. Ал шетел тәжірибесінде лингвистикалық зерттеулер үшін корпус базасы ертеден қолға алынып, түрлі математикалық әдістің көмегімен сөз және форма жиілігінің статистикасы анықталады. Шетелдік тәжірибеге негізделіп жасалған «Қазақ тілінің ұлттық корпусының подкорпустары» (qazcorporaz.kz) базасындағы корпусның іздеу механизмі сөзформа жиілігін зерттеуге аса қолайлы. [Qazcorporaz.kz](https://qazcorporaz.kz) сайтының алғашқы бетіндегі «Статистика» айдарында сөзформа жиілігін зерттеу үшін сілтеме ұсынылған. Оның көмегімен қажетті статистикалық ақпаратты автоматты түрде көшіріп алуға болады. Мақалада «Қазақ тілінің ұлттық корпусының подкорпустары» (qazcorporaz.kz) базасындағы жиілік статистика бойынша публицистикалық және ауызекі мәтіндеріндегі қазақ тілі етістіктерінің грамматикалық форма жиілігі талданады. Корпус базасындағы мәтіндерде ең жиі қолданылатын есімшелі, көсемшелі және қалып етістіктердің грамматикалық формалары, атап айтқанда, *деп, деген, емес, еді, болып, екен, болды, болған, алып, болады, келген, болса, жатқан, отыр, дейді, келеді, деді, болатын, жатыр, алған, келіп, жүрген, отырып, тұр, отырған, тұрған, жүріп, алды, тұратын, жүр, алатын, тұрып* етістіктерінің жиілігі өзара салыстырылады. Бұл етістіктердің мәтіндерде жиі қолданылу себептеріне семантикалық және контекстік талдау жасалады. *Де, е, бол, ал, кел, отыр, тұр, жүр, жат* көмекші етістіктерінің қазақ тіліндегі көпмағыналығы мен көпқызметтілігі талданады. Публицистикалық және ауызекі мәтіндерінде *деп, емес, болып, алып, келген, жатқан, отыр, тұр, жүрген* етістіктерінің тіркесімділік қабілетінің жоғары болатыны, олардың алуан түрлі синтаксистік қатынаста жұмсалып (анықтауыштық, пысықтауыштық, предикаттық), еркін, түйдекті тіркес және күрделі формалар (аналитикалық етістіктер) жасауға негіз болатын қабілеті жүйеленеді. Етістіктердің жиілік көрсеткіші кемімелі прогрессия және қарапайым есептеу жолымен шығарылды. Осы етістіктердің функционалдық қолданыс жиілігіне қатысты тұжырым мен қорытынды шығарылды.

Кілт сөздер: ұлттық корпус, мәтін, етістік, грамматикалық форма, жиілік көрсеткіші.

Кіріспе

Мақаламыздың зерттеу нысаны ретінде Ш. Шаяхметов атындағы «Тіл - Қазына» ұлттық ғылыми-практикалық орталығының «Қазақ тілінің ұлттық корпусының подкорпустары» (<https://qazcorporaz.kz>) дерекқорлары негіз болды. Корпус базасындағы газет материалдары және ауызекі сөйлеу мәтіндеріндегі етістіктердің грамматикалық формалары жиілігін анықтау барысында корпусның автоматтандырылған статистикасы басшылыққа алынды.

Осы стильдерге қатысты 13 миллион сөзформаның (2023 жылғы базалық деректер) негізінде жиі қолданылатын етістіктердің грамматикалық форма жиілігі қарастырылды.

Бұқаралық ақпарат құралдарының тілі ең өзекті әлеуметтік-мәдени және тілдік факторлардың айнасы іспеттес. Өйткені қоғамдағы жаһандану және цифрландыру процестерімен байланысты түрлі

* Хат-хабар үшін автор.e-mail: tancopan@gmail.com

құбылыстың, жаңалықтың атаулары ретінде жаңа сөздердің қоғамға енуі, ең алдымен, бұқаралық ақпарат құралдарында көрініс табады.

Сондықтан мақаламыздың мақсаты — публицистикалық және ауызекі сөйлеу тілінде жиі қолданылатын етістіктердің грамматикалық формалары жиілігін анықтау. Бұл аспект тіл білімінде әлі қарастырыла қоймаған мәселе болғандықтан, зерттеу өзекті болады. Корпус дерекқорындағы есімшелі, көсемшелі және қалып етістіктердің грамматикалық түрленуін, яғни етістіктердің жиілігін зерттеудің ғылыми және практикалық маңыздылығы болашақта бірқатар тіл мәселесін шешуге, атап айтқанда:

- қазақ тілі грамматикалық ережелерінің қаншалықты тұрақты, құбылмалы екенін анықтауға;
- қазақ тілінің синтаксистік құрылымын түсінуге;
- тілді өңдеуге;
- сөз қолданысының динамикалық өзгерістерін, стильдік ерекшеліктерін анықтауға және әлеуметтік-лингвистикалық вариация мен тілдің өзгеруін талдауға мүмкіндік береді.
- сондай-ақ машиналық тіл модельдеріне сөйлемдердің мағыналарын дұрыс түсінуге және генерациялауға мүмкіндік береді.

Зерттеудің әдістері мен материалдары

Зерттеуге алынған нысанға қол жеткізу үшін корпусстың іздеу механизміне жүгіндік. Корпусстың мақсаты қазақ тілін зерттеуге арналғандықтан, платформада ұсынылған корпусстың іздеу механизмдері сөзформа жиілігін зерттеуге аса қолайлы болды.

Qazcoqora.kz сайтының алғашқы бетіндегі «Статистика» айдарында корпус базасындағы барлық сөзформа жиілігін көшіріп алуға ыңғайлы сілтеме ұсынылған. Оның көмегімен қажетті статистикалық ақпаратты автоматты түрде көшіріп алуға болады. Корпусағы барлық сөзформа жиілігіне қатысты деректер арасында сәйкессіздік пен қайталануды болдырмас үшін автоматтандырылған тазалау скрипті қолданылған. Жалпы сөзқолданыстың жиілігі корпусстың жүйесіне қондырылған жиілік талдау бағдарламалық құралдары арқылы есептелген.

Сонымен қатар сөз жиілігін арнайы графикалық әдіспен жекелеген сөздерді енгізіп іздеуге де, мәтіндерден нақты іздеп, жалпы саны мен құжатта кездесу жиілігін де анықтауға арналған бағдарлама енгізілген.

Осындай корпусстың автоматтандырылған жүйесін пайдалана отырып, біз мақаламыздағы жиі қолданылатын етістіктердің грамматикалық формасының жиілік деректерін кемімелі прогрессия және қарапайым есептеу әдістерімен алдық. Ішінара семантикалық және контекстік әдістер де басшылыққа алынды. Корпуса сөздер жиілігінің реті ең көп және жиі қолданылатын сөзден басталады. Статистикалық әдіспен қатар салыстырмалы және талдау, қорытындылау әдістері де қолданылды. Лингвистикалық ұлттық корпус шетел тәжірибесінде әр қырынан зерттелген. Корпусы зерттеген шетелдік ғалымдардың еңбектеріне қысқаша шолу жасайтын болсақ, зерттеуші А. Kilgarrif корпусстардың ұқсастығы мен біркелкілігін өлшеу әдісімен анықтауға барған, грамматикалық және лексикалық ұқсастық арасындағы байланыстарға, сөз жиілігіне статистикалық эксперимент жасаған [1; 8-11]. Ал зерттеуші Н. Ваауен сөз жиілігінің статистикалық модельдерін зерттеген. Сөз жиілігін анықтайтын үш математикалық моделін: логнормал заңы (lognormal law), Гаусс-Пуассонның жалпыланған кері заңы (generalized inverse Gauss-Poisson law) және Zipf заңын (extended generalized Zipf's law) салыстырған [2; 348] Nick C. Ellis әлеуметтік лингвистикалық вариация мен тілдің өзгеру құбылысындағы жиіліктің рөлін зерттеген. Фонология, фонотактика, оқу, жазу емлесі, лексика, морфосинтаксис, шаблондық сөйлеу, тілді түсіну, грамматика, сөйлем құрылысын өңдеу кезіндегі жиіліктің әсерін қарастырған [3; 143].

Сондай-ақ шетелдік ғалымдар тарапынан көлемі үлкен корпусстарда жиі қолданылатын сөздерді анықтау барысында мәтіннің лексикалық негізіне грамматикалық мағына үстемелеп, мәтіндегі ойды байланыстыратын предлог, есімдіктер, шылауларға және зат есімдер, етістіктер, сын есімдер, үстеулер сияқты ерекше сөздердің орын алатыны біршама зерттелген [4; 23].

Кейбір зерттеу мақалаларында сөз жиілігі корпуссты талдаудың бастапқы нүктесі ретінде өте пайдалы болғанымен, оларды пайдаланудың белгілі проблемалары, яғни сөз жиілігін салыстырмалы бұрын, оларды бірыңғайландыру, жүйелеу маңызды екендігі тұжырымдалады. Заманауи тіл корпустарындағы сөз жиіліктері статистикалық әдістер арқылы салыстырылады [5; 81]. Зерттеуші S.T. Piantadosi сөздердің жиілігін статистикалық лингвистиканың негізгі зерттеу объектісі ретінде қарастыра отырып, сонымен қатар тілдік бірліктердің жиілігін коммуникативтік жүйенің

құрылымдық және функционалдық қасиеттерін анықтайтын негізгі сипаттамалардың бірі ретінде бағалайды, бұл сөздердің таралу заңдылықтарын зерттеуде олардың рөлін түсінуге мүмкіндік береді. [6; 11]. F. Nilsson ғылыми-көпшілік және ғылыми-зерттеу мақалаларындағы сөз жиілігін өзара салыстырған. Әрбіреуі 200 000 мыңнан астам сөзден тұратын екі жанрға сандық және сапалық зерттеу жүргізу үшін корпустық лингвистикалық талдау жасаған [7; 11].

Корпусты зерттеу мәселесіндегі заманауи зерттеулерге келетін болсақ, Қазақ тілінің ұлттық корпусының кіші корпусының (qazcorpora.kz) жиілік статистикасы деректері негізінде публицистикалық және ауызекі сөйлеу тілінде қолданылатын сөз таптары жиілігіне статистикалық талдау жасалған [8; 872]. Сондай-ақ корпус Qazcorpora.kz материалы негізінде қазақ тілінің эмоционалды бірліктерінің лексикалық, функционалдық ерекшеліктері талданған. Зерттеу жұмысының нәтижесінде қазақ тіліндегі эмоционалды бірліктер мен эмоционалды сөздер арасындағы айырмашылық анықталған [9; 62]. Келесі зерттеуде 2023 жылы БАҚ тілінің дискурсында ең жиі қолданылатын сөздер мен сөз тіркестері ұсынылып, оларға сандық талдау жасалады [10; 685].

Сонымен қатар қазақ-орыс параллельді А. Байтұрсынұлы мәтіндерінің корпусы негізінде аударма кезіндегі грамматикалық, стилистикалық өзгерістер талданған зерттеулер де бар [11; 763]. А. Байтұрсынұлының жинағына енген мәтіндер корпустық әдістер арқылы талданып, олардың сөз формаларының жиілігі корпус деректеріндегі сөз қолданылысымен салыстырылған, сонымен қатар тілдің динамикалық өзгерістеріне қатысты болжамдар жасалған [12; 16].

Қазақ тіл білімінде сөз жиілігі мәселесі лексикографиялық, әдістемелік тұрғыдан біршама зерттелген. Қазақ тілінде сөз жиілігіне қатысты «Қазақ тілінің жиілік сөздігі» (А.Қ. Жұбанов, А.Ә. Жаңабекова және т.б., 2016), «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» (2016), «Қазақ ауызша тіл мәтіндерінің жиілік сөздігі» (А.Қ. Жұбанов, А.Ә. Жаңабекова және т.б., 2020) жарық көрген. Осы бағытта «Қазіргі қазақ газеттері тілінің алфавитті-жиілік сөздігі» (Ә. Ахабаев, 1973) зерттеу еңбегін атауға болады.

Корпус материалының әдістемелік тиімділігі жайлы қазақстандық ғалымдардың: «Using corpora as authentic materials in the language learning classroom» (M. Zhambylkyzy, 2022), «Ұлттық корпустарға негізделген лингвистикалық зерттеулер жүргізу (қазақ, орыс, ағылшын тілі материалдары негізінде)» (К.Қ. Пірманова және т.б., 2022), «Применение корпусных методик в обучении иностранных студентов русскому языку и литературе» (С.Б. Бегалиева, Е.С. Шмакова, 2022) және т.б. еңбектері кездеседі.

Сонымен қатар «Индикативные формы глагола в казахском языке (формы прошедшего времени)» (Ж. Түймебаев, 1992) деп аталатын зерттеу еңбекті ерекше атап кетуге болады. Бұл еңбекте көркем шығарма, фольклор, ғылыми, публицистикалық әдебиеттер материалы негізінде етістіктің өткен шақ формаларының түрлі тілдік стильде қолданылу жиілігі мен сәйкестігі талданған.

«Сөйлеу етістіктерінің мағынасы мен қолданысы» тақырыбына (Г.О. Сыздықова, 2023) жазылған мақалада сөйлеу тілінде жиі қолданылатын *айт, сөйле, жеткіз* етістіктеріне семантикалық және контекстік талдау жасалған.

Сондай-ақ Қазақ тілінің ұлттық корпусының кіші корпусының (qazcorpora.kz) жиілік статистикасы деректері негізінде публицистикалық және ауызекі сөйлеу тілінде жиі қолданылатын сөз таптары, шылау және есімдіктер жиілігіне статистикалық талдау жасалған мақала (Omarova, S., Ospanova, D., Aitova, N., et al., 2025) жарық көрді.

Орыс тіл білімінде етістік формалары жиілігі мәселесі белсенді зерттелуде және түрлі аспектіде қарастырылады. Мәселен, «Семантическая структура глаголов с высокой частотностью» деген мақалада етістік лексикасының семантикасы мен жиі қолданысы арасындағы өзара байланыс зерттелген [13; 47].

«Частотность употребления иллокутивных глаголов в текстовых представлениях речевых актов» атты мақалада публицистикалық дискурста жиі қолданылатын етістіктердің қатары анықталып, теориялық талдау жасалған [14; 130].

Ал «Специфика употребления временных глагольных форм в английском языке при обучении устной диалогической речи» деген еңбекте ағылшын тілінде ауызша сөйлеуге үйретуге арналған тиімді оқулық жасау үшін етістіктің шақ формаларының қолданыс жиілігіне жасалған зерттеулер сипатталады [15; 24].

М.И. Грахольскаяның «Семантика английских глаголов разных частотных рангов» мақаласында ағылшын тіліндегі етістіктердің семантикалық және жиілік ерекшеліктерінің сәйкестіктеріне квантитативтік талдау жасалған [16; 12].

Сондай-ақ «Frequency of Verbal Forms and Language Standard» мақаласында Google Books корпусын пайдалану арқылы сөз түбірінде а/о дауыстылары өзгеріске ұшырайтын етістіктердің жиілігі зерттелген [17; 92].

Мақаланы жазу барысында Қазақ тілі ұлттық корпусы («Тіл – Қазына» ҰҒПО дайындаған) мәтіндеріндегі етістік формаларының жиілігін зерттеу үшін корпус базасындағы 2011 жылдан 2023 жылға дейінгі «Ана тілі» газеті, «Егемен Қазақстан» газеті, «Зан» газеті, «Түркістан» газеті, «Қазақ әдебиеті» газеті, «Айқын» газеті, «Ақиқат» журналы, «Астана ақшамы» газеті, «Жас алаш» және тағы басқа газет материалдарымен жұмыс жасалды. Бұл материалдар корпуста «Басылым бойынша статистика» деген айдармен ұсынылған.

1 - кесте

Етістіктердің грамматикалық форма жиілігі

| 1 | | 2 | | 3 | | 4 | | 5 | |
|----------|-----------|----------|-----------|---------|-----------|---------|-----------|---------|-----------|
| деп | 23769 жк1 | емес | 13607 жк3 | болып | 11051 жк5 | алып | 7196 жк9 | келген | 6200 жк11 |
| деген | 17272 жк2 | еді | 11110 жк4 | болды | 7932 жк7 | алған | 4124 жк19 | келеді | 5273 жк15 |
| дейді | 5427 жк14 | екен | 8697 жк6 | болған | 7861 жк8 | алды | 2341 жк27 | келіп | 3930 жк20 |
| деді | 5188 жк16 | | | болады | 6992 жк10 | алатын | 1496 жк29 | келетін | 1207 жк31 |
| | | | | болса | 6047 жк12 | | | | |
| | | | | болатын | 4477 жк17 | | | | |
| б | | 7 | | 8 | | 9 | | | |
| жатқан | 6047 жк12 | отыр | 5498 жк13 | жүрген | 3689 жк21 | тұр | 3525 жк23 | | |
| жатыр | 4474 жк18 | отырып | 3534 жк22 | жүріп | 2362 жк26 | тұрған | 3006 жк25 | | |
| жатып | 621 | отырған | 3468 жк24 | жүр | 1693 жк29 | тұратын | 1699 жк28 | | |
| жатырмыз | 508 | отырмыз | 1010 жк32 | жүретін | 645 | тұрып | 1488 жк30 | | |
| жатамын | 461 | отыратын | 512 | жүрміз | 343 | тұрмыз | 83 | | |

Ескерту: 1. (жк1, жк2, жк3 және тағы басқа осындай белгілер) — жиілік көрсеткішінің реттілік саны

Корпустық статистика негізінде 9 етістіктің грамматикалық формаларының жиілігі өзара салыстырылған. Онда 1010-нан 23769-ға дейінгі 9 етістіктің грамматикалық форма жиілігі реттілік саны бойынша көрсетілген. Ең жиі қолданылатыны — көсемшелі формадағы *деп* көмекші етістігі. Оның жиілігі — 23769.

Жалпы *де* етістігі *айт*, *сөйле* етістіктерімен мағыналас, синонимдес. Сондықтан да осы мағына ерекшелігі бойынша оны сөйлеу етістіктері тобына жатқызамыз.

Етістіктің семантикалық топтарының ішінде ең жиі қолданылатыны — сөйлеу етістіктері. Сөйлеу етістіктері адам үшін ең маңызды қажеттіліктің бірі — сөйлеуді жүзеге асырады. Г.О. Сыздықова өзінің «Сөйлеу етістіктерінің мағынасы мен қолданысы» мақаласында *айт*, *сөйле*, *жеткіз* етістіктерінің сөйлеу тілінде жиі қолданылатын ерекшеліктерін тоқталған [18; 52].

Алайда корпустың сөзформа жиілігі статистикасына үңілетін болсақ,

айт етістігінің «айт+ып» көсемшелі формасының жиілік көрсеткіші — 3623, «айтты» жедел өткен шақ формасыныкі — 1798 екен. Яғни екі сөзформасының жиілік көрсеткіштерін салыстыратын болсақ, көсемшелі формасы 49,6 % жиі қолданылады. Ал «сөйле» етістігінің сөйле+n және сөйле+di грамматикалық формаларының жиілік көрсеткіші өте төмен. Корпус статистикасында «сөйлеп — 393», «сөйледі — 131» көрсеткіштеріне ие.

Сондықтан публицистикалық стиль мен сөйлеу тілінде *деп* көсемшелі формасының *айт*, *сөйле* етістіктеріне қарағанда жиі қолданысы оның көпмағыналылығы мен көпқызметтілігіне байланысты деуге болады.

Де етістігінің көпмағыналылығы мен көпқызметтілігі оның көмекшілік қызметіне де қатысты. Оның осындай ерекшелігі ешбір көмекші етістікке тең келе алмайды. Мысалы: Уай, не *деп* кеттіңдер?! (С. Сағынтай «Майқұдық оқиғасы»).

Қазақ тілінде *деп* көмекші етістігі төл сөз бен автор сөзін байланыстырушы қызметімен ерекшеленеді. Мысалы: Шешеме: «Енді не істейміз?», – *деп* едім, ол иығын көтеріп: «Ауладағы кішкене жерден басқа ештеңеміз жоқ қой енді», – деді (Д. Кішібеков «Қазақтың бүкіл тұрмыс-тіршілігі — философия»).

Корпус базасынан алынған алғашқы синтаксистік конструкцияда *де* көмекші етістігінің көсемшелі формасының сөйлем ішінде бір рет емес, екі рет қолданылғанын байқаймыз:

1. Мен: «Ауылда жалғыз інім ғана бер еді», – деп едім, «жоқ, академияға жібереміз», – деп, айтқанынан қайтпады.

Бір сөйлемнің ішінде *деп көмекші етістігінің* 3 рет қайталануы да кездеседі. Мысалы: Шамасы, Мира күйеуі кешірім сұрайды, «кетпеші» *деп*, аяғыма жығылады *деп* ойлаған болуы керек, «енді не істеймін?» *деп* бөгеліңкіреп қалды (А. Есімбай «Жарым»). Осы мысалдағы *деп көмекші етістігіндегі* күрделі тіркестер пысықтауыштық қатынасқа құрылған.

Төл сөзді автор сөзімен байланыстырғанда, *де* көмекші етістігі түрлі тұлғада (*деп, дейді, деген, деді* т.б.) қолданылатыны мәлім. Дегенмен публицистикалық мәтінде *де+n* грамматикалық формасының жиілігі басқаларына қарағанда жоғары. Көсемшелі формасының жиілігі (*деп*) – 23 769 болса, *есімшелі формасы (деген)* — 17 272, *ауыспалы осы шақ формасы (дейді)* — 5428, *жедел өткен шақ формасы (деді)* — 5188 жиілікті құрайды.

Де көмекші етістігі бірде есімше, бірде көсемше тұлғалы етістіктердің жұрнақтары арқылы түрленіп, синтаксистік тіркестің бағыныңқы сыңарында жұмсалады да, құрылымына қарай күрделі еркін тіркесті құрайды. А. Ысқақов: «Ең алдымен бұл форма хабарланылмақшы, ойға қатысты сөзді я сөйлемді түгел үйіріп, айтылмақшы, я аталмақшы амал-әрекеттің (көбінесе, әрине, баяндауыштың) мақсаты, себеп-салдары, нұсқауы сияктандырып көрсету үшін қолданылады. Мұндағы *деп формасы* өзінің негізгі мағынасын түгел я жартылай сақтап отырады» [19; 269], – дейді .

Сонымен қатар *деп көмекші етістігі* есім сөзбен, бұйрық райлы, қалау райлы етістікпен де түйдекті тіркес құрайды. Әрі аталған көмекші етістік өзіне қатысты, өзі тіркескен есім сөздердің, етістіктердің синтаксистік қызметін өзгертеді [20; 93].

Төмендегі мысалдарға назар аударсақ, *деп көмекші етістігі* зат есіммен, көсемшелі және жедел өткен шақ формалы қосарланған етістіктермен, міндеттілік мәнін білдіруші модаль сөздермен қолданылып түйдекті тіркес жасап, басыңқы сыңарда жұмсалып, анықтауыштық және пысықтауыштық қатынаста тұр.

1. Менің ойымша, біздің қазақ *деп аталатын* ұлтымыз, бүкіл түркітөктес халықтар ғана емес, бүкіл әлем елдерінің ішінде ең озық ойшыл халық (Д. Кішібеков «Қазақтың бүкіл тұрмыс-тіршілігі — философия»).

2. Сонда, үш күннің ішінде 60 мыңдай адам *келіп-кетті деп* шамалауға болады (Қ.М. Табеев. «Дерегі көп, керегі жоқ желтоқсан»).

3. *Күтіп аламын деп* әуре болма (Н. Ораз «Шыбын жан көкке ұшқанда»).

4. Оған Сәруарды *ауылдасым, бауырым деп* таныстырды (Н. Ораз «Шыбын жан көкке ұшқанда»).

Келесі *е көмекші етістік* формаларының емес, еді, екен қатарындағы емес болымсыз грамматикалық формасы соңғыларымен салыстырғанда қолданысы жоғары, яғни 13607 жиілікті құрайды.

Болымсыз есімше тұлғасындағы *емес* көмекші етістігінің публицистикалық контексте тіркесімділік функциясы жоғары, мәселен: *мүмкін емес, тиіс емес, керек емес, бар емес пе, ғана емес (қана емес), аз емес, өзінің емес, өзгелер үшін емес, мен емес, бірінші рет емес, бір емес, бауырың емес, олардай емес, бос емес, алдым деп емес, емес пе?, оқитындығымнан емес, нәтижесіз емес, да(те) емес, Шадияров емес* және т.б.

Осы тіркесімдер ішінде *емес* көмекші етістігі публицистикалық мәтіндерде көп жағдайда модаль сөздермен тіркесіп қолданылады.

Келесі *еді* формасының да жиілігі жоғары, ол — 11110. Бұл етістік тілде күрделі формалар жасауға қатысып, есім сөздермен де, етістіктермен де тіркесе береді [21; 50]. Мәселен: *айтатын еді, студия еді, жаз еді, шыққан еді, емдер еді, қандай еді, қайсысы еді, айтып еді, ала алмас еді, азаятын еді, болатын еді* және т.б.

Алайда публицистикалық мәтіндерде *еді* формасының модаль сөздермен де жиі қолданылатын ерекшелігі бар. Мысалы: *бар еді, жоқ еді, керек еді, сияқты еді, мүмкін емес еді*.

Бұл қатардағы *екен* формасының жиілік көрсеткіші — 8697. Корпус мәтіндеріндегі қолданысы, мәселен: *көңіліне түйсе екен, тынышыса екен, жатыр екен, демейді екен, жемейді екен, арыла алмайды екен, береді екен, қадам басқан екен, үлес қосқан екен және т.б.* Публицистикалық мәтінде белсенді қолданылады. Бұл грамматикалық формалардың жиі қолданылуы тек қазақ тіліне ғана тән емес, сонымен қатар қырым татар тілінде де жиі қолданылатын ерекшелікке ие.

Екен формасы жайлы ғалым А.М. Меметов былай деген: «Этимологически *екен* восходит к форме причастия кыпчакского типа, образованной от древнего глагола *эр* («быть, становиться») посредством известного аффикса *-кен* (один из фонетических вариантов аффикса *-гъан*), в результате образования новой формы *екен* в современных языках придаёт ей значение «был», «оказывается, (кажется) был» [22; 219].

Қазіргі қырым татар тілін зерттеген ғалым Л.И. Бекирова (2016) тұжырымына сәйкес, өткен шақты білдіретін «екен» көмекші етістігімен жасалатын формалар жиі қолданылса да, «еді» көмекші етістігімен салыстырғанда «екен» құрылымының жиілігі төмен болып келеді [23; 24].

Корпустық статистикаға сүйенетін болсақ, қазақ тілінің публицистикалық және сөйлеу стильдерінде *еді* формасының жиілік қолданысымен салыстырғанда *екен* формасы 22 % төмен қолданылады екен (1-кесте, Етістіктердің грамматикалық форма жиілігі).

Қимылдың аяқталғандығын білдіретін және құранды етістік жасауға белсенді қатысатын етістіктердің бірі — *бол көмекші етістігі*. Ол мәнді көмекші етістіктердің қатарына жатады, яғни жеке тұрғанда лексикалық мағынаға ие болады да, көмекші етістік қызметін атқарғанда өзінің лексикалық мағынасынан айырылып, тіркескен сөзге қимылдық мағына үстейді. Оның *-ып, -ды, -ған* формалары ең жиі қолданылатын формаларға жатады, ал *-а+ды* (ауыспалы келер шақ), *-са* (шартты рай түрі), *-атын* (есімшелі түрі) формаларының жиілігі төмендеу (*1-кесте*).

Бұл грамматикалық формалар публицистикалық және сөйлеу тілінде жиі қолданылуымен қатар, көркем шығармалар тілінде де жиілік қабілеті жоғары.

Етістік формаларының жиілігі жөніндегі тұжырымдарға сәйкес ғалым Ж. Түймебаев былай деген: «Түрлі стиль мен жанрлар мәтіндерінде етістіктің шақ формаларының қолданысын зерделей отыра, көркем шығармаларда етістіктің өткен шақ формалары жиі қолданылатынын байқадық. Егер көркем шығармаларда *-ды* формасы белсенді қолданылса, ал ғылыми, ғылыми-көпшілік мәтіндерде *-ған* формасы бірінші орында тұрады» [24; 105].

Бол көмекші етістігінің «болып» көсемшелі формасының *-ды, -ған, -а+ды, -са, -атын* грамматикалық формаларымен салыстырғанда жиілігі өте жоғары. *Бол* етістігінің көсемшелі формасы жедел өткен шақ формасынан 27,3 %, есімшелі формасына қарағанда 29 %, ауыспалы келер шақ формасынан 37 %, шартты рай формасынан 45,3 %, есімшенің *-атын* формасынан 59 % жиі қолданылады.

Ол публицистикалық стильде есім, модаль, еліктеуіш сөздермен және түрлі формадағы етістіктермен тіркесіп қолданылады. Мысалы: *жапондар ел болып, ақылдасып алуым керек болып тұр, пора-пора болып ақты, жолым болып түсіп кеттім, бәрі риза болып жатты, Атырауда болып келдім, аппақ болып киініп жүретін, түп-түзу болып тұратын, ерекше жақын болып, аудармашы болып жұмыс істедім, төрағасы болып барғанмын, экспонат болып сақталып тұр, «әжесінің қызы» болып өстім, Нұр-Сұлтан қаласы болып өзгерді, шашу болып шашылған, кім көрінгенге жем болып және т.б.*

Жалпы *бол* етістігінің жоғарыда аталған грамматикалық формаларының белсенділігі публицистикада өте жоғары.

Келесі «ал» негізгі түбір етістігінің де көсемшелі формасы қолданыс жиілігі — 7196. «Алған» формасының қолданыс жиілігі 43 % болса, «алып» формасынікі — 57 %. «Ал» етістігінің *ал+ды, ал+атын* формалары есімшелі және көсемшелі формаларымен салыстырғанда сирек қолданылады. «Алды» етістігі негізгі және көмекші етістік қызметінде жұмсалатыны белгілі. Мәселен: *телефон алды, жеті-сегіз адам алды, трубкасын алды, өзі алды, тиынды алды; ішіп алды, шақырып алды, қарсы алды, қолға алды, айналдырып алды, отырып алды, жайғасып алды, созып алды, шырмап алды және т.б.* Бұл етістік публицистикада көбінесе *-ып алды* құрылымында жиі кездеседі, қолданыс жиілігі — 33 %.

Қазақ тілінде «ал» етістігінің дағдылы өткен шақ есімшелі формасы (*ал+атын*) бұрын болған әрекеттің бірнеше қайталанып, дағдылы қалыпта орындалып тұрғанын білдіретіні белгілі, мәселен: *сабақ алатын, жаздырып алатын, сатып алатын, тігіп алатын, тауып та алатын және т.б.*

Осы *-атын* формалы есімшелі етістіктің мағынасында келер шақты білдіру қабілеті де бар, мысалы, публицистикалық контексте: *гонорар алатын болды, алатын секілді* және т.б.

Дегенмен *-атын* жұрнақты есімшенің шақтық мағынадан алшақтап, атрибуттық мағынада жұмсалып, анықтауыштың қызметін атқаратын қасиетке де ие: *таласа алатын мазмұн, сөйлесе алатын жағдай, айта алатын пенде, қамтамасыз ете алатын нәрсе* және т.б.

Сонымен «ал» етістігі жиілігі жағынан тоғызыншы орында болса, оның *-ған* формасы 18 орында, *-ды* формасы 26, *ал -алатын* формасы 29 орында.

Келесі «кел» етістігінің бұрынғы өткен шақ есімшелі *-ген* формасы *-e+di, -in және -етін* формаларымен салыстырмалы түрде жиі қолданылады. Публицистикадағы *кел+ген* формасының тіркесімділігі бірде атрибуттық бірде предикаттық қолданысынан көрінеді, мәселен: *экскурсияға келген қонақтар, өмірге келген буын, ауылдан келген дәмнен, реті келген жерде сезінбей келген адаммын, айтқысы келген ойын, бойшаң келген жігіт; көргісі келген, дүниеге келген, талай келген, келген еді, барып келген* және т.б. Корпус мәтіндерінде «келген» формасының атрибуттық қызметте жұмсалуды көбірек орын алады.

Корпус статистикасындағы жиілік көрсеткіші бойынша «кел» етістігінің есімшелі формасы (келген) 11 орында, ауыспалы келер шақ формасы (келеді) 15 орында, көсемшелі формасы (келіп) 19 орында және дағдылы өткен шақ есімшелі формасы (келетін) 31 орында.

Енді қазақ тіліндегі төрт қалып етістіктерінің ішінен «отыр» етістігі көп қолданылатыны анықталды. Деректер (1-кесте) бойынша *отыр* етістігінің жиілігі — 5498, *жатыр* — 4474, *тұр* — 3525, *жүр* — 1693.

«Отыр» етістігінің корпус базасындағы қолданысына тоқталып көрейік. Мысалы: *қамтып отыр, жұмыссыз отыр, алдымда отыр, үйде отыр*.

Алайда «жатыр, отыр, тұр, жүр» етістіктерімен салыстырғанда «жатқан» формасының жиілігі жоғары, яғни 6047 құрайды.

Сонымен қатар «жатыр» формасына қарағанда «*жатқан*» етістігінің, «отырып, отырған, отырмыз» етістіктеріне қарағанда «*отыр*» қалып етістігінің, «тұрған, тұратын, тұрып» етістіктеріне қарағанда «*тұр*» формасының, «жүріп, жүр» етістік формаларына қарағанда «*жүрген*» формасы жиі қолданылады.

Қорытындылай келе, *де, бол, ал* етістіктерінің көсемшелі формалары, *кел, жат, жүр* етістіктерінің бұрынғы өткен шақ есімшелі формалары, *е* көмекші етістігінің болымсыз есімше тұлғасы, сондай-ақ *отыр, тұр* қалып етістіктерінің нөлдік формалары жиі қолданылатыны анықталды. Демек, *де+n, e+мес, бол+ып, ал+ып, кел+ген, жат+қан, отыр, тұр, жүр+ген* етістіктерінің публицистикалық мәтіндерде тіркесімділік қабілетінің жоғары болатыны, олардың алуан түрлі синтаксистік қатынаста жұмсалып (анықтауыштық, пысықтауыштық, предикаттық), еркін, түйдекті тіркес және күрделі формалар (аналитикалық етістіктер) жасауға негіз болатын қабілетінде деуге болады.

Осындай бағытта басқа да етістіктердің грамматикалық форма жиілігін зерттеуге болады.

Корпус базасы жылдан-жылға жетілдіріліп отыратындықтан, базадағы сөз қолданыс саны ұлғайған сайын сөзформа жиілігінің көрсеткіші де өзгеріп отырады.

Корпус публицистикалық және көркем мәтіндерді қамтыса да, тілдік вариация мен өзгерісті терең зерттеу үшін оның көлемін арнайы ауызша, тарихи мәтіндерден тұратын кіші корпустармен кеңейту қажет.

Зерттеушілер мен тілші мамандардың қажырлы еңбегінің арқасында күннен-күнге дамып, қоры байып келе жатқан Қазақ тілінің ұлттық корпусы тілдің мүмкіндігін танытатын, оны түсіну мен қолдануды жеңілдететін және зерттеуді жеделдететін қазақ тілінің басты цифрлы базасына айналуға [25; 142].

Қорытынды

Сөз және грамматикалық форма жиілігі экстралингвистикалық факторға байланысты құбылып отырады. Өйткені сөйлеушінің жазу және сөйлеу кезіндегі сөзді және грамматикалық формаларды қолдану сипаты тұрақты болмайды, уақыт өткен сайын өзгереді. Бұндай зерттеу жұмыстары контекстегі сөзформаның құбылуы немесе тұрақтануы мәселесіне қатысты ғылыми зерттеулерге жол ашады.

Қазақ тілін арнайы зерттеу үшін және әдістемелік-дидактикалық мақсатта пайдалану үшін жасалған «Қазақ тілінің ұлттық корпусы подкорпустарының» (<https://qazcorpora.kz>) жиілік

статистикасы грамматикалық ереженің құбылмалылығын, грамматикалық формалардың динамикалық өзгерісін, әлеуметтік-лингвистикалық вариация және тағы басқа келелі мәселені зерттеуге мүмкіндік беретіні белгілі.

Сондай-ақ етістіктердің грамматикалық форма жиілігін анықтау әдістемелік мақсатты оңтайлы шешуге мүмкіндік береді.

Осы мақалада зерттеу нысанына алынған 9 етістіктің (*де, е, бол, ал, кел, жат, отыр, жүр, тұр*) грамматикалық форма жиілігі талданды. Жалпы 41 грамматикалық форманың жиілігі өзара салыстырылды. Оның ішінде 34 грамматикалық форманың жиілік көрсеткіші жоғары, ал 7 грамматикалық форманың (*жат+ын, жатыр+мыз, жат+атын, отыр+атын, жүр+етін, жүр+міз, тұр+мыз*) жиілігі төмен екендігі анықталды.

Етістіктің грамматикалық формаларының жиілік статистикасын өзара салыстыру барысында көсемшелі, есімшелі, болымсыз және жедел өткен шақ формаларының (*де+n, де+ген, е+мес, е+ді, бол+ын*) жиі қолданылатыны функционалдық және синтаксистік қызметіне байланысты талданды.

Зерттеу жұмысы «Қазақ тілін және технологиялық прогресті қолдау үшін үлкен тілдік модельді (LLM) құру» (ИРН: BR24993001) жобасы аясында орындалды.

Әдебиеттер тізімі

- 1 Kilgarriff A. Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora [Electronic resource] / A. Kilgarriff // Proceedings of the Fifth Workshop on Very Large Corpora. — 1997. — P. 231–245. — Access mode: <https://aclanthology.org/W97-0122>
- 2 Baayen H. Statistical models for word frequency distributions: A linguistic evaluation / H. Baayen // Computers and the Humanities — 1992. — Vol. 26, No. 5. — P. 347–363. DOI: <https://doi.org/10.1007/BF00136980>
- 3 Ellis N.C. Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition / N.C. Ellis // Studies in Second Language Acquisition. — 2002. — Vol. 24, No 2. — P. 143–188. DOI: <https://doi.org/10.1017/S0272263102002024>
- 4 Scott M. Textual patterns: Key words and corpus analysis in language education / M. Scott, C. Tribble. — Amsterdam: John Benjamins Publishing Company, 2006. — 167 p.
- 5 Baron A. Word frequency and key word statistics in historical corpus linguistics / A. Baron, P. Rayson, D. Archer // Anglistik: International Journal of English Studies. — 2009. — Vol. 20, No. 1. — P. 41–67.
- 6 Piantadosi S.T. Zipf's word frequency law in natural language: a critical review and future directions / S.T. Piantadosi // Psychonomic Bulletin & Review. — 2014. — Vol. 21, No. 5. — P. 1112–1130. DOI: <https://doi.org/10.3758/s13423-014-0585-6>
- 7 Nilsson F. A comparative analysis of word use in popular science and research articles in the natural sciences: A corpus linguistic investigation / F. Nilsson. — Västerås: Mälardalen University, 2019. — 50 p.
- 8 Omarova S. A Corpus Approach in Language Discovery: A Word Frequency Analysis Based on the Corpus Outcomes in Kazakh / S. Omarova, D. Ospanova, N. Aitova et al. // Forum for Linguistic Studies. — 2025. — Vol. 7, No 2. — P. 869–881. DOI: <https://doi.org/10.30564/fls.v7i2.8317>
- 9 Aitova N. Verb-based Emotive Structures in the Linguistic Corpus Base / N. Aitova, D. Ospanova // Bulletin of Toraigrov University. Philological Series. — 2024. — No 1. — P. 55–69.
- 10 Ormanova A.B. A corpus-based approach in vocabulary research: Defining the word of the year 2023 in Kazakh / A.B. Ormanova, M.L. Anafinova, D.Zh. Ospanova // Theory and Practice in Language Studies. — 2024. — Vol. 14, No 11. DOI: <https://tpls.academypublication.com/index.php/tpls/article/view/9825>
- 11 Baishukurova, G. A Corpus-Based Analysis: Kazakh-Russian Grammatical Transformations in Akhmet Baitursynuly's Translations / G. Baishukurova, A. Irgebayeva, N. Aitova et al. // Forum for Linguistic Studies. — 2025. — Vol. 7, No 1. — P. 753–766. DOI: <https://doi.org/10.30564/fls.v7i1.7993>
- 12 Аитова Н.Н. Ахмет Байтұрсынұлы шығармаларын корпусдық зерттеу («Қырық мысал», «Маса» жинақтары негізінде) / Н.Н. Аитова, Г.Ж. Байшукурова, А.Б. Иргебаева // ҚазҰУ хабаршысы. Филология сериясы. — 2024. — Т. 29, № 3 (115). — Б. 14–23.
- 13 Плотникова С.В. Семантическая структура глаголов с высокой частотностью / С.В. Плотникова // Функциональная семантика слова. — Екатеринбург: Уральский педуниверситет, 1994. — С. 46–51.
- 14 Назаретян Н.В. Частотность употребления иллокутивных глаголов в текстовых представлениях речевых актов / Н.В. Назаретян // Филологические науки. Вопросы теории и практики. — Тамбов: Грамота, 2014. — № 12 (42). — С. 128–131.
- 15 Митрофаненко Л.М. Специфика употребления временных глагольных форм в английском языке при обучении устной диалогической речи / Л.М. Митрофаненко, М.А. Павлюковец, О.Г. Цыганская. — Ставрополь: Северо-Кавказский федеральный университет, 2019. — С. 11–28.

- 16 Грахольская М.И. Семантика английских глаголов разных частотных рангов / М.И. Грахольская // Германские языки. Филологический аспект. — 2018. — № 12 (44). — С. 12–25.
- 17 Galeev T. Frequency of Verbal Forms and Language Standard / T. Galeev, E. Habibulina // Journal of History Culture and Art Research. — 2017. — Vol. 6, No. 5. — P. 90–98. DOI: <http://dx.doi.org/10.7596/taksad.v6i5.1250>
- 18 Сыздықова Г.О. Сөйлеу етістіктерінің мағынасы мен қолданысы / Г.О. Сыздықова // Л.Н. Гумилев атындағы ЕҰУ Хабаршысы. Филология сериясы. — 2023. — № 1 (142). — Б. 52–57.
- 19 Ысқақов А. Қазіргі қазақ тілі / А. Ысқақов. — Алматы: Ана тілі, 1991. — 384 б.
- 20 Қияқова Р.Ж. Сөйлем мүшесі қызметіндегі «Де» тұлғалы түйдекті тіркес / Р.Ж. Қияқова // Қазақстан Республикасы Ұлттық ғылым академиясы Хабаршысы. — 2013. — Б. 92–95.
- 21 Жолшаева М.С. Қазақ тіліндегі құранды етістіктердің лексика-грамматикалық ерекшеліктері / М.С. Жолшаева // Tiltanym. — 2022. — № 1 (85). — Б. 44–54.
- 22 Меметов, А.М. Крымскотатарский язык: учебное пособие: [в 2 т.] / А.М. Меметов, К.М. Мусаев. — Симферополь: Крымучпедгиз, 2003. — Т. 1-2. — 288 с.
- 23 Бекирова Л.И. Прошедшее время глагола на — имя + екен в современном крымскотатарском языке / Л.И. Бекирова // Ученые записки Крымского федерального университета имени В.И. Вернадского. Серия Филологические науки. — 2016. — Т. 2 (68). — № 3. — С. 20–24.
- 24 Түймебаев Ж.Қ. Индикативные формы глагола в казахском языке (формы прошедшего времени) / Ж.Қ. Түймебаев. — Алма-Ата: Ана тілі, 1992. — 116 с.
- 25 Момынова Б.Қ. Қазақ тілінің ұлттық корпусындағы етістіктердің лексика-семантикалық белгіленім әзірлемесі: әлемдік тәжірибе, классификациялау, корпуста белгілеу / Б.Қ. Момынова, М.А. Имангазина, У.Г. Анесова // Абылай хан атындағы ҚазХҚ және ӨТУ Хабаршысы. Филология ғылымдары сериясы. — 2022. — № 3 (66). — Б. 128–146.

С.К. Ыдырыс, Л.Ж. Бейсенбаева, Д.Ж. Оспанова

Частота грамматических форм в текстах корпуса (на основании <https://qazcorpora.kz>)

В данной статье рассматривается частотность грамматических форм глаголов в рамках вопроса о том, насколько эффективно мы можем использовать корпусные материалы для изучения грамматических аспектов казахского языка. Использование текстов из корпусной базы для лингвистических исследований имеет большое значение. Анализ корпусных текстов позволяет выявлять изменения и развитие лексического слоя, синтаксических конструкций и грамматических форм в языковой системе. Так как словарный запас любого языка изменчив и подвержен трансформациям, важно ежегодно отслеживать, какими новыми словами и словоформами он пополняется, каков уровень устойчивости языка к заимствованиям из других языков, в какой форме он принимает новые словоформы. Для этого корпусная база должна регулярно пополняться и совершенствоваться. Без подобных исследований направление развития нашего языка может стать неясным. В зарубежной практике корпусные базы давно используются для лингвистических исследований, с применением различных математических методов проводится статистический анализ частотности слов и форм. Подкорпусы Национального корпуса казахского языка (qazcorpora.kz), созданные на основе зарубежного опыта, имеют удобные механизмы поиска, что делает их особенно полезными для изучения частотности словоформ. В разделе «Статистика» на главной странице сайта qazcorpora.kz предоставлена ссылка, с помощью которой можно автоматически получить необходимую статистическую информацию. В статье рассматривается частотность грамматических форм казахских глаголов в публицистических и разговорных текстах на основе статистических данных *Подкорпуса Национального корпуса казахского языка* (qazcorpora.kz). Особое внимание уделяется сравнительному анализу наиболее часто употребляемых грамматических форм деепричастий, причастий и глаголов состояния. В частности, исследуется частотность таких глаголов, как: *деп, деген, емес, еді, болып, екен, болды, болған, алып, болады, келген, болса, жатқан, отыр, дейді, келеді, деді, болатын, жатыр, алған, келіп, жүрген, отырып, тұр, отырған, тұрған, жүріп, алды, тұратын, жүр, алатын, тұрып*. Проведён семантический и контекстуальный анализ причин высокой частотности указанных глагольных форм в текстах корпуса. Анализируется многозначность и многофункциональность вспомогательных глаголов *де, е, бол, ал, кел, жүр, отыр, тұр, жат*, а также их роль в формировании устойчивых словосочетаний и аналитических форм в публицистическом и разговорном стиле. Выявлены особенности сочетаемости глаголов *деп, емес, болып, алып, келген, жатқан, отыр, тұр, жүрген* в устойчивых словосочетаниях и проанализированы их функции в создании различных синтаксических отношений (атрибутивных, обстоятельственных, предикативных). Систематизированы их основные функции в образовании свободных и сложных словосочетаний, а также аналитических форм глагола. Сделаны выводы о частоте и функциональной значимости данных глагольных форм в современных казахских текстах.

Ключевые слова: национальный корпус, текст, глагол, грамматическая форма, частотный индекс.

S.K. Ydyrys, L.Zh. Beysenbayeva, D.Zh. Ospanova

Frequency of grammatical forms in corpus texts (based on <https://qazcorpora.kz>)

This article examines the frequency of grammatical forms of verbs within the framework of the question of how effectively we can utilize corpus materials to study the grammatical aspects of the Kazakh language. Using texts from the corpus base for linguistic research is highly important. For this reason, the corpus base must be regularly updated and improved. Without such studies, the direction of language development becomes unclear. In international practice, corpus databases have long been employed in linguistic research, using various mathematical methods to analyze the frequency of words and forms. The subcorpora of the National Corpus of the Kazakh Language (qazcorpora.kz), developed based on international experience, offer convenient search mechanisms that are especially useful for studying the frequency of word forms. On the main page of the qazcorpora.kz website, under the “Statistics” section, a link is provided that allows users to automatically obtain the necessary statistical information. The article examines the frequency of grammatical forms of Kazakh verbs in journalistic and spoken texts based on frequency statistics of the Subcorpus of the National Corpus of the Kazakh language (qazcorpora.kz). Special attention is paid to the comparative analysis of the most frequently used grammatical forms of adverbs, participles, and verbs of state. In particular, the frequency of verbs such as: *dep*, *degen*, *emes*, *edi*, *bolyp*, *eken*, *boldy*, *bolgan*, *alyp*, *bolady*, *kelgen*, *bolsa*, *zhatkan*, *otyr*, *deidi*, *keledi*, *dedi*, *bolatyn*, *zhatyr*, *alghan*, *kelip*, *jurgen*, *otyrp*, *tur*, *otyrghan*, *turgan*, *zhurip*, *aldy*, *turatyn*, *zhur*, *alaty*, *turyp*. A semantic and contextual analysis of the reasons for the high frequency of these verb forms in the corpus texts is carried out. The article analyzes the ambiguity and versatility of auxiliary verbs *de*, *e*, *bol*, *al*, *kel*, *zhur*, *otyr*, *tur*, *zhat*, as well as their role in the formation of stable phrases and analytical forms in journalistic and colloquial style. The features of the compatibility of the verbs *dep*, *emes*, *bolyp*, *alyp*, *kelgen*, *zhatkan*, *otyr*, *tur*, *jurgen* in stable phrases are revealed and their functions in creating various syntactic relations (attributive, circumstantial, predicative) are analyzed. Their main functions in the formation of free and complex phrases, as well as analytical verb forms, are systematized. Conclusions are drawn about the frequency and functional significance of these verb forms in modern Kazakh texts.

Keywords: national corpus, text, verb, grammatical form, frequency index

References

- 1 Kilgarriff, A. (1997). Using Word Frequency Lists to Measure Corpus Homogeneity and Similarity between Corpora. *Proceedings of the Fifth Workshop on Very Large Corpora*, 231–245. Retrieved from <https://aclanthology.org/W97-0122>
- 2 Baayen, H. (1992). Statistical models for word frequency distributions: A linguistic evaluation. *Computers and the Humanities*, 26, 5, 347–363. DOI: <https://doi.org/10.1007/BF00136980>
- 3 Ellis, N.C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in Second Language Acquisition*, 24, 2, 143–188. DOI: <https://doi.org/10.1017/S0272263102002024>
- 4 Scott, M., & Tribble, C. (2006). *Textual patterns: Key words and corpus analysis in language education*. Amsterdam: John Benjamins Publishing Company.
- 5 Baron, A., Rayson, P., & Archer, D. (2009). Word frequency and key word statistics in historical corpus linguistics. *Anglistik: International Journal of English Studies*, 20, 1, 41–67.
- 6 Piantadosi, S.T. (2014). Zipf’s word frequency law in natural language: a critical review and future directions. *Psychonomic Bulletin & Review*, 21, 5, 1112–1130. DOI: <https://doi.org/10.3758/s13423-014-0585-6>
- 7 Nilsson, F. (2019). *A comparative analysis of word use in popular science and research articles in the natural sciences: A corpus linguistic investigation*. Mälardalen: Mälardalen University.
- 8 Omarova, S., Ospanova, D., & Aitova, N., et al. (2025). A Corpus Approach in Language Discovery: A Word Frequency Analysis Based on the Corpus Outcomes in Kazakh. *Forum for Linguistic Studies*, 7, 2, 869–881. DOI: <https://doi.org/10.30564/fls.v7i2.8317>
- 9 Aitova, N., & Ospanova, D. (2024). Verb-based Emotive Structures in the Linguistic Corpus Base. *Bulletin of Toraigyrov University. Philological Series*, 1, 55–69.
- 10 Ormanova, A.B., Anafinova, M.L., & Ospanova, D.Zh. (2024). A corpus-based approach in vocabulary research: Defining the word of the year 2023 in Kazakh. *Theory and Practice in Language Studies*, 14, 11. DOI: <https://tpls.academypublication.com/index.php/tpls/article/view/9825>
- 11 Baishukurova, G., Irgebayeva, A., & Aitova, N., et al. (2025). A Corpus-Based Analysis: Kazakh-Russian Grammatical Transformations in Akhmet Baitursynuly’s Translations. *Forum for Linguistic Studies*, 7, 1, 753–766. DOI: <https://doi.org/10.30564/fls.v7i1.7993>
- 12 Aitova, N.N., Baishukurova, G.Zh., & Irgebayeva, A.B. (2024). Akhmet Baitursynuly shygarmalaryn korpustyq zertteu («Qyryq mysal», «Masa» zhinaqtary negizinde) [Corpus study of Ahmet Baitursynuly’s works (based on the collections “Forty Prov-

erbs”, “Masa”)]. *Qazaq Ultyq Universitetinin Khabarshysy. Filologiya seriiasy — Bulletin of the Kazakh National University. Philological Series*, 29, 3, 14–23 [in Kazakh].

13 Plotnikova, S.V. (1994). *Semanticheskaiia struktura glagolov s vysokoi chastotnostiu* [Semantic structure of high-frequency verbs]. Ekaterinburg: Uralskiy Peduniversitet [in Russian].

14 Nazaretyan, N.V. (2014). Chastotnost upotrebleniia illokutivnykh glagolov v tekstovykh predstavleniakh rechevykh aktov [Frequency of use of illocutionary verbs in textual representations of speech acts]. *Filologicheskie nauki. Voprosy teorii i praktiki — Philological Sciences. Issues of Theory and Practice*, 12, 42, 128–131. Tambov: Gramota [in Russian].

15 Mitrofanenko, L.M., Pavlyukovets, M.A., & Tsyganskaya, O.G. (2019). *Spetsifika upotrebleniia vremennykh glagolnykh form v angliiskom yazyke pri obuchenii ustnoi dialogicheskoi rechi* [Specificity of the use of tense verb forms in English when teaching oral dialogic speech]. Stavropol: Severo-Kavkazskii federalnyi universitet [in Russian].

16 Grakholskaya, M.I. (2018). Semantika angliiskikh glagolov raznykh chastotnykh rangov [Semantics of English verbs of different frequency ranks]. *Germanskii yazyki. Filologicheskii aspekt — Germanic Languages. Philological Aspect*, 12, 44, 12–25 [in Russian].

17 Galeev, T., & Habibulina, E. (2017). Frequency of Verbal Forms and Language Standard. *Journal of History Culture and Art Research*, 6, 5, 90–98. DOI: <http://dx.doi.org/10.7596/taksad.v6i5.1250>

18 Syzdykova, G.O. (2023). Soileu etistikterinin magynasy men qoldanysy [Meaning and use of speech verbs]. *L.N. Gumilev atyndagy Eurazia Ultyq Universitetinin Khabarshysy — Bulletin of the L.N. Gumilyov Eurasian National University. Philologi Series*, 1, 142, 52–57 [in Kazakh].

19 Ysraqov, A. (1991). *Qazirgi qazaq tili* [Modern Kazakh language]. Almaty: Ana tili [in Kazakh].

20 Kiaqova, R.Zh. (2013). Soilem mushesi qyzmetindegi «De» tulgalı tuidekti tirkes [The phrase with the form “De” in the function of a sentence member]. *Qazaqstan Respublikasy Ultyq Gylım Akademiiasy Khabarshysy — Bulletin of the National Academy of Sciences of the Republic of Kazakhstan*, 92–95 [in Kazakh].

21 Zholshaeva, M.S. (2022). Qazaq tilindegi qurandy etistikterdin leksika-grammatikalıq erekshelikteri [Lexical and grammatical features of compound verbs in the Kazakh language]. *Tiltanyım — Linguistics*, 1, 85, 44–54 [in Kazakh].

22 Memetov, A.M., & Musaev, K.M. (2003). *Krymskotatarskii yazyk* [Crimean Tatar language]. (Vols. 1-2; Vol. 2). Simferopol: Krymchpedgiz [in Russian].

23 Bekirova, L.I. (2016). Proshedshee vremia glagola na — imia + eken v sovremennom krymskotatarskom yazyke [Past tense of the verb on — name + eken in the modern Crimean Tatar language]. *Uchenye zapiski Krymskogo federalnogo universiteta imeni V.I. Vernadskogo — Scientific Notes of the V.I. Vernadsky Crimean Federal University*, 2(68), 3, 20–24 [in Russian].

24 Tuymebaev, Zh.Q. (1992). *Indikativnye formy glagola v kazakhskom yazyke (formy proshedshego vremeni)* [Indicative forms of the verb in the Kazakh language (past tense forms)]. Almaty: Ana-tili [in Russian].

25 Momynova, B.K., Imangazina, M.A., & Anesova, U.G. (2022). Qazaq tilinin ulttyq korpusyndagy etistikterdin leksika-semantikalıq belgilenim azirlemesi [Development of Lexica-Semantic Marking of Verbs in the National Kazakh Corpus of the Kazakh Language: World Experience, Classification, Markup in the Corpus]. *Abylai khan atyndagy Qazaq Khalyqaralyq Qatynastar zhane Alem Tilderi Universitetinin Khabarshysy. Filologia gylımdary seriiasy — Bulletin of Kazakh Ablai Khan University of International Relations and World Languages. Series “Philological Sciences”*, 3, 66, 128–146 [in Kazakh].

Information about the authors

Ydyrys, Sofiya Kazhiakparovna — Candidate of Philological Sciences, Shaisultan Shayakhmetov National Scientific and Practical Center “Til – Qazyna”, Astana, Kazakhstan. E-mail: omarovask@mail.ru; orcid.org/0009-0009-0819-0543

Beisenbayeva, Lyazzat Zhumabekovna (corresponding author) — PhD, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. E-mail: tancopan@gmail.ru; orcid.org/0000-0001-5932-9263

Ospanova, Dana Zhanarbekkyzy — PhD, L.N. Gumilyov Eurasian National University, Astana, Kazakhstan. E-mail: dana.zhanarbekkyzy@gmail.com; orcid.org/0000-0001-8901-8060