

Қазақстан Республикасы Білім және Ғылым министрлігі  
Ғылым комитеті  
А.Байтұрсынұлы атындағы Тіл білімі институты

А.Қ. ЖҰБАНОВ, А.Ә.ЖАҢАБЕКОВА

**ҚАЗАҚ ЖАЗУЫН ЛАТЫН ҚАРПІНЕ  
КӨШІРУДІҢ СТАТИСТИКАЛЫҚ  
НЕГІЗДЕРІ**

*Алматы, 2018*

ӘОЖ 811.512.122  
КБЖ 81.2 Қаз 2  
Ж 80

**А.Байтұрсынұлы атындағы Тіл білімі институтының  
Ғалымдар кеңесі ұсынған**

Монография «Латынграфикалы жаңа ұлттық әліпби негізінде қазақ жазуын жаңғыртудың ғылыми-лингвистикалық базасын әзірлеу» атты мақсатты-бағдарламалық жоба аясында дайындалды

**Пікір жазғандар:**

**Е.З.Қажыбек** – филология ғылымдарының докторы, профессор,  
**С.Қ.Құлманов** – филология ғылымдарының кандидаты

**Ж 80 Жұбанов А.Қ., Жаңабекова А.Ә.**

Қазақ жазуын латын қарпіне көшірудің статистикалық негіздері.  
– Алматы, «А.Байтұрсынұлы атындағы Тіл білімі институты» баспасы,  
2018 ж. – 160 б.

**ISBN 978-601-7293-62-8**

Оқырман назарына ұсынылып отырған «Қазақ жазуын латын қарпіне көшірудің статистикалық негіздері» атты зерттеу жұмысы «Қазақ графемаларының жиілік сөздігіне» статистикалық талдау жасауға арналған. Бұл еңбекте қазақ тіліндегі әріп, әріп тіркесі, буын, түрленім қосымша, түрленім форма, сөз графемаларының әрқайсысына жиілік сөздік бойынша әр қырынан статистикалық мәліметтер берілген. Сонымен қатар осы тілдік бірліктердің қазақ статистикасының зерттелу тарихындағы деректері қоса келтірілген. Статистикалық талдау нәтижесінде алынған мәліметтер қазақ тіл білімінің лингвостатистика саласының теориялық мәселелерін толықтыратын болады және жалпы тілдік зерттеулер үшін нақты статистикалық ақпараттар береді. Әсіресе, қазақ жазуын латын графикасына көшіру ісінде, атап айтқанда, емле, орфография ережелерін жазуда, пернетақта жасауда фактологиялық дереккөз болып табылады.

ӘОЖ 811.512.122  
КБЖ 81.2 Қаз 2

ISBN 978-601-7293-62-8

© Жұбанов А.Қ., Жаңабекова А.Ә., 2018

## АЛҒЫ СӨЗ

Жазба тілі сөйлеу тілі сияқты практикалық және теориялық тұрғыда аса маңызды зерттеу нысаны екені мәлім. Мысалы, ақпаратты автоматты түрде өңдеу мен шартты белгілермен таңбалау (кодтау) кезеңінде сол тілдің графемдерінің статистикалық сипаттамаларын білу керек болады. Әсіресе, қазіргі таңда қажеттілік тудырып отырған қазақ тілін латын қарпіне көшіруге қатысты мәселені шешудің ғылыми-практикалық алғышарттарын жасау ісінде маңызды болып табылады.

Қазақ тіл білімінде ғылыми зерттеулерде статистикалық әдіс жиі қолданылғанмен, лингвостатистика мәселесімен арнайы айналысқан ғалымдар саны көп емес. Тіл білімі саласының сан ұғымымен тығыз байланысты болып, математика саласымен астасып жатқандықтан, бұл мәселеге математик ғалымдар алдымен қызығушылық танытты. Қазақ статистикасының негізін қалаған белгілі қазақ математигі Қалдыбай Бектаев XX ғасырдың 70-жылдары А.Байтұрсынұлы атындағы Тіл білімі институтында Компьютерлік лингвистика бөлімін ашып, әртүрлі жиілік сөздіктер жасау мәселесін көтерді. Қазақ тіл білімінде кезінде 70-жылдары компьютерлік технология дами қоймаса да, қазақ тіл білімінде алғашқы жиілік сөздікті құрастырушылар сөздердің әртүрлі мәтіндердегі кездесу жиілігін ЭЕМ көмегімен анықтады.

Түркі тілдері бойынша алғаш боп жиілік сөздіктер алуды бастаған қазақ тілші ғалымдары әртүрлі стильдер мен белгілі бір автор шығармалары бойынша жиілік сөздіктер алуды сәтті орындап шықты. Түркі тілдері бойынша статистикалық зерттеулер жүргізіліп, жиілік сөздіктердің 70-жылдары алғаш қазақ тілінде жарық көруі сол кезеңдегі үлкен жаңалық еді. Бұл еңбектер күні бүгінге дейін маңызын жоғалтқан емес. Қазақ тілінің статистика саласында еңбек еткен азғантай топтың шығарған жиілік сөздіктері, статистикалық зерттеулері – құнды зерттеулердің өміршендігінің дәлелі.

Осы жиілік сөздіктер негізінде ғалымдар «Қазақ тексінің статистикасы» атты еңбекте өз зерттеулеріне статистикалық талдаулар жасады. Бұдан кейін 2016 жылы А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімі белгілі статист-ғалым профессор А.Жұбановтың жетекшілігімен бұрын-соңды жарық көрген әртүрлі стильдегі жиілік сөздіктердің басын қосып, біртұтас «Қазақ тілінің жиілік сөздігін» құрастырды. Осы сөздіктің кіріспе бөлімінде де жиілік мәліметтерге сөз жиілігі, сөз табы жиілігі т.б. қырларынан статистикалық талдаулар жасалды. Сондай-ақ өзге ұлт өкілдеріне мемлекеттік тілді оқыту мақсатында лексикалық минимумдар жасау үшін Білім және ғылым министрі Е.Сағадиевтің тапсырмасы бойынша Институт А.Фазылжанованың басқаруымен «Жалпы білім берудегі қазақ



## КІРІСПЕ

Ғылыми-техникалық прогресс, электронды-есептеу техниканың дамуы, нақты әдіс-тәсілдердің түрлі білімдер салаларында қолданыс табуы тіл біліміне, соның ішінде түркітануға да өз ықпалын тигізді.

Өткен ғасырдың өзінде үндіеуропа тіл білімінде кең тараған статистикалық әдіс түркітануда, соның ішінде қазақ тілінде тек XX ғасырдың 60-жылдарынан бастап жүйелі және ғылыми тұрғыда қолданыла бастады. Бірақ бұдан бұрын да аталған әдіс ішінара қолданыс тауып, тілдің әр деңгейіндегі бірліктердің сандық және пайыздық арасалмағын анықтаумен шектеліп келген болатын.

Түркі тіліндегі мәтіндерді статистикалық тәсілмен зерттеу – бұл тілдің ішкі табиғатына, яғни агглютинативті құрылымына негізделеді. Мамандардың пайымдауынша, кітаби және сөйлеу тіліндегі түрлі типті сөздердің әр ыңғайдағы қолдану жиілігін анықтайтын мұндай тәсіл тілдің құрылымдық-грамматикалық, кейде стильдік-семантикалық айырым белгілерін ажырата отырып, осы стильдерге тән қасиеттерді анықтауға мүмкіндік жасайды. Сонымен бірге грамматикалық категорияларды сандық тәсілмен талдау әдеби тілдің барлық аясындағы функционалдық салмағын көрсетуде де маңызды.

Белгілі филолог-статист Р.Г.Пиотровский түркі тілдерін сандық деңгейде зерттеудің негізі олардың ішкі құрылымдық табиғатына тән екендігін айта келе, бұл тілдерге үндіеуропа тіліне тән екпінді ассимиляция мен редукциялар және осы сипатты қасиеттер жат екенін баса көрсетеді. Бұл жағдай тілдік деректердің жай кезде қиындық туғызатын кейбір тұстарын математикалық әдіспен тіркеу мен тануда аса маңызды деп саналады.

Тілдік мәселелерге қолданылатын математикалық әдісті интуитивті түрде тұжырымдалған, сол сияқты толық шешімі жоқ мәселелердің логикалық тұжырымдалуына бағынатын және алгоритмдік шешімі бар бір немесе бірнеше қарапайым математикалық есептермен ауыстыру деп ұғыну қажет. Тілдік мәселені осылайша математикалық жолмен шешу – лингвистикалық нысанның математикалық моделіне (үлгісіне) көшуді қажет етеді. Тілдің аса күрделі түрдегі сан алуан параметрлі көп қабатты жүйе екенін ескерсек, онда ықтималдық теория, математикалық статистика, ақпарат теориясы және т.б. ғылыми салалар әдістерін қолданудың мүмкіндігі кең екені байқалады.

Тіл ерекшелігі мәтіндердің жанрлық (стильдік) түрлеріне қарай да айырым табатыны белгілі. Осындай айырымдарды қазақ тілінің әртүрлі жанларына қатысты статистикалық жолмен тағайындауды қазақстандық статистика тобының ғалымдары 70–80-жылдары өз мақсаты етіп қойды.

Мысалы, Қ.Молдабеков пен Б.Е.Қалыбековтер балалар әдебиеті мен бастауыш сыныптарының оқулықтарын зерттеуді өз міндеттеріне алса, А.Р.Зекенова – М.Әуезовтің драмалық шығармалар тілін, А.Ахабаев – газет тілін, А.Белботаев – ғылыми-техникалық стиль, математикалық стиль тілдерін өз зерттеу нысандары етіп алып, алғашқы ғылыми нәтижелерге қол жеткізді.

Қазақ тілінің фонетика саласы да статистика әдісінің назарынан тыс қалған жоқ. Фонетиканың дәстүрлі (есту) және құралды-тәжірибелік зерттеу әдісі бойынша фонологиялық және графикалық жүйелерге қатысты мәселелерін ықтималды-статистикалық әдіс-тәсілдермен шешу тиімді деп танылды.

Қазақ тіл білімінің қолданбалы саласының тың проблемалық мәселесі – қазақ мәтінінің әр әріпке (дыбысқа) шаққандағы ақпараттық өлшемін, яғни *энтропияны* анықтау. Осындай зерттеулер белгілі ғалым Қ.Б.Бектаев пен оның шәкірті Д.А.Байтанаеваның қазақ мәтініне қатысты ақпараттық-статистикалық ғылыми ізденістері арқылы іске асырыла бастады. Табиғи тілдің белгілер жүйесі үш түрлі деңгейде – *синтактикалық, семантикалық* және *прагматикалық* аспектілерде қарастырылатыны мәлім. Мұндағы синтактика деңгейі әртүрлі белгілер жүйесінің синтаксисін, яғни сөз мағынасынан тыс, белгілердің тіркесімдік құрылымын және олардың туындау ережелерімен қатар белгілер құрылымының ішкі қасиеттерін қарастырады. Ал белгілер жүйесінің сыртқы құрылымдық қасиеттерімен семантика мен прагматика айналысады. Дәлірек айтсақ, егер *семантика* белгілер жүйесін мағыналық көзқарас, яғни мәтін мазмұны тұрғысынан танып-білумен шұғылданса, *прагматика* – белгілер жүйесінің қабылдаушыға қатынасын анықтау мәселелерімен айналысады.

Қазіргі кезде ақпарат аясын теориялық және технологиялық тұрғыда зерттеудің синтактикалық деңгейі жақсы дамып келеді. Бұған К.Шеннонның символдар жүйелілігі мен комбинаторикасын қарастыратын статистикалық байланыс теориясы негіз болады. Бұл теория бойынша табиғи тілде жазылған мәтін белгілі бір код құрайтын дискретті белгілердің тізбегі ретінде қарастырылады. Ал мұндай мәтінді қабылдау кезінде дискретті белгілердің қай-қайсысының да пайда болуы белгісіздіктің шамасымен, яғни энтропиямен өлшенеді. Қ.Б.Бектаев пен Д.А.Байтанаеваның ғылыми ізденістерінде қазақ мәтінінің ең кіші бірлігі – әріпке түсетін энтропияның орта шамасын және қазақ тілінің толық және стильдік (жанрлық) айырымдарын анықтайтын жазба мәтіндеріндегі энтропия мен артық ақпарат шамаларын анықтау негізгі мақсат ретінде қойылды. Осымен қатар қазақ тілі бойынша анықталған осындай сипаттамаларды синтактикалық деңгейде үндіеуропа тілдері бойынша алынған белгілі деректермен салыстыра зерттеулер де іске асты. Қазақ мәтініндегі әрбір әріпке түсетін энтропияның ең үлкен (макси-

мум) шамасын анықтау, әліпби әріптерін қабылдаудың ықтималдықтары тең және өзара тәуелсіз дәрежеде деген ұйғарымға негізделетіні айқын.

Статистикалық деректерді өңдеуде компьютерлік технологияны пайдалану, тілдік мәселелердің шешімін тез арада табу, математикалық статистика әдістерін басқа ғылымдар тәсілдерімен (ақпараттық теория, математикалық логика және т.б.) ұштастыру қазіргі кездегі отандық және шетелдік тіл білімінің даму бағытына сай келеді.

Түркі мәтіндерін статистикалық тәсілдермен зерттеу, ғалымдардың пайымдауынша, бұл тілдердің ішкі табиғатына, яғни олардың жалғамалық (агглютинативтік) құрылымына негізделеді.

Тіл зерттеу тәжірибесінде сандық зерттеулердің қажеттігін математикалық лингвистика мамандарымен қатар дәстүрлі тіл білімінің аса көрнекті өкілдері де мойындап келеді. Олардың пікірінше, жазба әдебиет пен сөйлеу тіліндегі түрлі сөздердің әр ыңғайдағы қолданылу жиілігін анықтау тәсілі тілдің құрылымды-грамматикалық, кейде стильді-семантикалық айырымдарын ажыратумен қатар стильдерге тән қасиеттерді де анықтауға көмектеседі. Сондай-ақ қазақ тілін кирилл әріптерінен латын әліпбиіне көшіруге (ауыстыруға) қатысты зерттеулерде статистикалық әдіс-тәсілдерді қолдану, атап айтқанда, *әріп, әріптікес, қосымша, түрленім форма, буын, сөз, сөзформа* және т.б. графемалардың жиілігіне статистикалық талдау жасау оң нәтижеге әкелетіні сөзсіз.

## 1 ҚАЗАҚ ӘРІПТЕРІ МЕН ӘРІП ТІРКЕСТЕРІНІҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ

### 1.1 Қазақ әріптері мен әріп тіркестерінің жиілік сөздігін құрастырудың қажеттілік негіздемесі

Әліпби мәселесін шешуде әріптер мен қазақ сөздеріндегі, мәтіндеріндегі әріп тіркестерінің жиілігін білудің қажеттігі дау тудырмайды. Әріптер мен әріп тіркестерінің жиіліктерін анықтау қазақ әліпбиін латын қарпіне көшірудегі бірнеше теориялық-практикалық бағыттағы мәселелерді шешуге септігін тигізеді.

*Біріншіден*, әріп, әріп тіркестерінің жиілігі мәтінді қолмен теруде жылдамдықты арттыру үшін латын қаріпті қазақ пернетақтасына әріптерді тиімді орналастыруда қажет болса, *екіншіден*, қазақ жазуын реформалаудағы орфографиялық емле ережелерін шешу барысында жасалатын зерттеулер мен лингвистикалық әзірленімдерге (разработка) фактологиялық сандық деректер үшін, *үшіншіден*, кирилл қаріпті мәтіндерді латын қарпіне автоматты аударма жасаудағы программаға қажетті емле ережелерін жасау барысында жиілік бойынша реттеу, сонымен қатар нысанға алынған әріп тіркестері туралы сандық мәліметтер беру және нысанға алынған әріп тіркестерін мысалмен дәйектейтін сөздерді қазақ тілінің мәтіндер корпусынан іздестіріп, көптеген мысал сөздерді табу сияқты мәселелерді шешуде қажет.

### 1.2 Қазақ әріптерінің жиілік сөздігін құрастыру тарихы

Қазақ тіл білімінде әріптерді статистикалық тұрғыдан зерттеудің тарихы 1973 жылы жарық көрген «Қазақ тексінің статистикасы» атты еңбектен басталады деуге болады. А.Қ.Жұбановтың (мақалада А.Х. Джубанов) «К вопросу о графемной статистике казахского текста», Қ.Бектаев пен С.Ризаевтың «Диалектілердің бір тілге жатуының критерийі жайында», А.Исеналиеваның «Фонемалардың салыстырмалы жиілігін анықтаушы факторлар», Э.Жұбанованың «Қазақ тілінің әріп және дыбыс құрамының жалпы статистикасы» атты еңбектерін атап айтуға болады.

Қ.Бектаев пен С.Ризаевтың «Диалектілердің бір тілге жатуының критерийі жайында» атты еңбекте түркі диалектологиясына статистикалық талдаулар жүргізуге арналған. Авторлар түркі тілдерінің ішінде өзбек тілінің говорларында қолданылатын фонемалардың жиілігін анықтаған және қазақ тіліндегі фонема жиіліктерімен салыстырған. Бір тілдегі говордағы фонема



жиіліктерімен қатар, өзбек және қазақ тіліндегі фонемалардың жиіліктерін салыстырып, қазақ тілінде жасалымына (жақтың қатысына) қарай жоғары-орта дауысты дыбыстардың, ал өзбек тілінде жақтың қатысына қарай ашық дауысты дыбыстардың, өзбек тілінде тіл ортасы және тіл алды дыбыстардың, ал қазақ тілінде тіл арты дауысты дыбыстардың жиі қолданылатынын, сондай-ақ қазақ тілінде ұяң, үнді және тілалды дауыссыз дыбыстардың жиі қолданылатынын анықтайды [1, 646 б.]. Сонымен бірге қазақ тілінде еріндіктер, ал өзбек тілінде езуліктер жиірек қолданылады дегенді айтады. Сонымен қатар авторлардың тұжырымдауынша, дыбыстардың қолданылу жиілігі өзбек тіліндегі сөздерді белгілі бір говорына жатқызудың критерийлерінің бірі болып табылады.

Осы мақалада дыбыс топтарының ғана емес, жекелеген фонемалардың да жиілігі беріледі. Авторлар қазақ тіліндегі фонемалардың жиілігін анықтап, өзбек тілімен салыстыру үшін қазақ ертегілерінен 20 707 дыбыс қолданыстан (фонемоупотребление) тұратын мәтін алған. Осы көлемдегі мәтінде қазақ дыбыстарының жиілігі берілген. **1-кесте.**

**1-кесте. Ертегі мәтіні бойынша қазақ әріптерінің жиілігі**

Қазақ әліпбиі	Абсолютті жиілік	Қазақ әліпбиі	Абсолютті жиілік
А	12 235	Ұ	2834
Е	7976	О	2173
Н	7165	К	2041
Ы	6306	Ү	2013
І	6166	Ң	1835
Т	4978	Ғ	1704
Р	4944	З	1560
Д	4934	Ш	1560
С	3978	Ж	1444
Л	3853	Ө	1424
П	3592	У	937
М	3018	Г	927
Б	2994	Ә	719
Қ	2892	Х (һ)	14
Й	2882	И	9

Ертегілер мәтінінде *в, ф, ц, ч, ь, ъ* әріптері кездеспеген.



**3-кесте. Ертегі және ғылыми проза мәтіні бойынша қазақ тіліндегі дауыссыз дыбыстардың жиілігі**

Әріп	Ертегі	Ғылыми проза	Әріп	Ертегі	Ғылыми проза
Н	7,17	6,25	Ж	1,44	1,72
Т	4,98	6,06	З	1,56	1,70
Л	3,85	5,03	П	3,59	1,53
Р	4,94	5,01	У	0,94	1,48
Д	4,93	4,72	Ш	1,56	0,99
С	3,98	4,28	Г	0,93	0,34
Қ	2,89	3,35	Ф	0,00	0,11
Б	2,99	3,05	Х	0,27	0,09
Й	2,88	3,04	В	0,00	0,07
М	3,02	2,61	Ч	0,00	0,03
Ғ	1,70	2,36	Ц	0,00	0,03
К	2,04	2,28	Һ	0,01	0,00
Ң	1,83	1,87	Щ	0,00	0,03

Берілген фонемалардың статистикасының жанрлық ерекшеліктеріне қарай жоғарыдағы А.Исенгельдинаның айтқан пайымдауларынан басқа да түрлі бағыттағы зерттеулер барысында әртүрлі теориялық тұжырымдар шығаруға болады.

Біздің зерттеу нысанымызға қажетті мәлімет дауысты дыбыстардан ең жиі кездесетіні – А фонемасы екендігі. Ал Е, Ы, І фонемалары барабар жиілікпен екінші орында тұрады. О, Ө, Ұ, Ү фонемалары барабар жиілікпен үшінші орында, ал И, У дауыстылары сирек кездесетінін көреміз.

Дауыссыздар бойынша ең жиі кездесетіні – Н фонемасы. Екінші орында – барабар жиілікпен Т, Л, Р, Д, С, М фонемалары. Үшінші орында – Ғ, К, Ң, Ж, З, П, Ш, У фонемалары, ал Г – сиректеу, Ф, Х, В, Ч, Ц, Һ Щ өте сирек, тіпті ертегілерде Х-дан басқасы кездеспейді.

Әріп жиілігіне қатысты тағы да жанрлық ерекшелікті сипаттайтын еңбек – Э.Жұбанованың «Қазақ тілінің әріп және дыбыс құрамының жалпы статистикасы» атты мақаласы.

Э.Жұбанованың статистикалық зерттеу жұмысының нысаны ретінде А.Байтұрсынұлы атындағы Тіл білімі институтында 2004 жылы жарық көрген «Қазақ тілінің орфоэпиялық сөздігінің» мәтіні алынған. Екі бағанадан тұратын бұл сөздіктің сол жақ бағанасында 75 мыңнан аса сөз бен сөз тіркестерінің орфографиялық жазылу нормасы (жызылу үлгісі) сақта-



11	і	28161	440720	0,04356	0,68170
12	б	20162	460882	0,03119	0,71289
13	д	19851	480733	0,03071	0,74359
14	к	17983	498716	0,02782	0,77141
15	м	16203	514919	0,02506	0,79647
16	ш	16174	531093	0,02502	0,82149
17	ғ	12976	544069	0,02007	0,84156
18	й	12430	556499	0,01923	0,86079
19	ж	11500	567999	0,01779	0,87857
20	о	10759	578758	0,01664	0,89522
21	з	10722	589480	0,01658	0,91180
22	п	8192	597672	0,01267	0,92447
23	ұ	7989	605661	0,01236	0,93683
24	ң	7072	612733	0,01094	0,94777
25	г	6996	619729	0,01082	0,95859
26	ү	6944	626673	0,01074	0,96933
27	ө	6309	632982	0,00976	0,97909
28	и	5817	638799	0,00900	0,98809
29	ә	3819	642618	0,00591	0,99399
30	я	1913	644531	0,00296	0,99695
31	ю	1177	645708	0,00182	0,99877
32	х	361	646069	0,00056	0,99933
33	һ	125	646194	0,00019	0,99953
34	в	121	646315	0,00019	0,99971
35	ф	50	646365	0,00008	0,99979
36	щ	44	646409	0,00007	0,99986
37	э	38	646447	0,00006	0,99992
38	ь	27	646474	0,00004	0,99996
39	ц	19	646493	0,00003	0,99999
40	ъ	1	646494	0,00001	1

Осы сөздіктің ортаңғы зонасына (2-топ), яғни әріптердің қолдану дәрежесін орта шама деп шартты қабылдау негізінде әріптердің мәтіндегі абсолютті жиілігі **1000**-нан көп, бірақ **10000**-нан аз болып келетін ( $1000 < F_i < 10000$ ) аралығы, яғни сөздіктегі қатар сандары **22**-ден **31**-ге тең **10** әріп алынған. Бұл әріптер қайталана қолдану негізінде барлық орфографиялық мәтіннің **53138** әріптік көлемді немесе **8,7 %** камтиды.

Сөздіктің үшінші – ең төмен жиілікті зонасына рет сандары **32** мен **40** аралығына тең **9** әріп шартты түрде алынған. Бұл әріптер негізінен кірме сөздерде қолданыс табатын әріптер қатарына жатады және барлық мәтіннің тек **786** әріптік көлемін немесе **0,12** пайызын ғана қамтыды.

Бірінші зонаға, яғни жоғары жиілікті дыбыстар тобына (шартты түрде) ең алғашқы орналасқан **25** қатар санын қамтитын дыбыстар таңбалары жатқызылған. Ол барлық дыбыстар қолданысының он мыңнан көп, бірақ жүз мыңнан аз болып келетін аралығы ( $10000 < F_i < 100000$ ). Бұл зона **672787** дыбыстық таңбалар көлемін немесе барлық мәтіннің **96,45%** қамтиды.

Екінші зона ретінде дыбыс қолданыстарының жүзден көп, бірақ он мыңнан аз болып келетін аралығы ( $100 < F_i < 10000$ ) шартты түрде саналған. Бұл қатар саны **26**-дан **31**-ге тең аралықтағы **6** дыбыс таңбасы **23641** дыбыстық айтылу мәтінін немесе барлық дыбыстық мәтін көлемінің **3,51%** қамтиды.

Үшінші зонаға, яғни дыбыстардың сирек қолданысына сәйкес келетін зона – абсолютті жиіліктері бір мен жүз сандары аралығындағы **9** дыбысты қамтиды. Бұл зона кейбір кірме сөздердегі дыбыстар негізінде пайда болып, барлық айтылу мәтіннің **0,12%** ғана қамтиды [3.].

А.Жұбановтың «К вопросу о графемной статистике казахского текста» атты еңбегінде әріптердің статистикасы жайлы мәліметтер берілген [4.]. Ғалым қазақ тілінің үш түрлі стилі (жанр) мәтіндеріндегі әріптердің жиілік сөздігін құрастырып, жанрлық статистикалық ерекшеліктерін салыстырған. Олар:

1) М.О.Әуезовтің «Абай жолы» романы мәтіні (**51290** сөзқолданыс немесе **280812** әріп);

2) математика пәніне арналған мектеп оқулықтарының мәтіні (**19467** сөзқолданыс немесе **130388** әріп);

3) қазақ тілінің екі томдық түсіндірме сөздігінің мәтіні (**17585** сөзқолданыс немесе **116317** әріп).

А.Қ.Жұбанов тәжірибе түрінде алынған стильдер мәтіндерінің тең көлемді болмауының себебі, олар толық түрдегі шығармалар болуымен түсіндіріледі. Сондықтан мәтіндер ішіндегі тілдік бірліктердің (әріптердің) сандық сипаттары жайлы мәлімет алу үшін олардың қатынастық жиіліктері (немесе пайыздық салмақтары) арқылы пайымдалғандығын ескертеді. Мақалада қазақ тілінің үш стилі бойынша компьютер көмегімен түзілген қазақ әріптерінің жиілік сөздігінен үзінді алынып, жиі кездесетін он әріптің жиілігі бойынша статистикалық талдау жасалған. **5-кесте.**

**5-кесте. Қазақ тілінің үш стилі (жанр) бойынша түзілген қазақ әріптерінің жиілік сөздігінен үзінді**

Стиль-дердегі әріптің алатын орны	Әріптердің мәтінді қамту пайызы (%)								
	Роман			Математика			Сөздік		
	Әріп аты	Әріптің %	Жиынтық %	Әріп аты	Әріптің %	Жиынтық %	Әріп аты	Әріптің %	Жиынтық %
1	А	13,2	13,2	А	11,0	11,0	А	13,0	13,0
2	Е-Ё	8,1	21,3	Е-Ё	8,3	19,3	Т	7,1	20,1
3	Ы	7,5	28,8	Н	8,2	27,5	У	7,0	27,1
4	Н	7,1	35,9	Ы	7,4	34,9	Е-Ё	6,3	33,4
5	І	6,0	41,9	Д	5,4	40,3	Ы	6,1	39,5
6	Т	5,0	46,9	І	5,3	45,6	Р	6,0	45,5
7	Р	4,8	51,7	Р	5,3	50,9	Л	6,0	51,5
8	Д	4,6	56,3	Т	5,2	56,1	С	4,5	56
9	Л	4,2	60,5	Л	5,1	61,2	Қ	4,1	60,1
10	С	4,1	64,6	С	4,4%	65,6	Н	4,1	64,2
Қосындысы:			65%			66%			64%

Кестеде ең жиі кездесетін әріптер олардың жиіліктерінің кему тәртібімен орналасқан. Кестеде он әріптің әрбіреуінің және олардың жиынтықтарының мәтінді қамту пайыздық салмағы көрсетілген. Кестедегі мәліметтерге сүйенсек, қазақ әліпбиіндегі **40** әріптің тек **10-ы** ғана әр стиль бойынша алынған мәтіндердің **64-66** пайыздарын қамтитынын байқауға болады. Ал көркем әдебиет (роман) стилі мен ғылыми-техникалық (математика) стилінде ең жиі қолданыс табатын: **а, е-ё, ы, н** төрт әріп (үш дауысты, бір дауыссыз) барлық мәтіннің **35-36** пайызын қамтиды екен. Бұл жерде әмбебап дауыстылар қатарына жататын **4-ші і** дауысты дыбысы аталған әр стильде 5-ші және 6-шы орындарға (іргелес орындарға) ие болып тұрғанын байқаймыз.

**5-кестедегі** сөздік мәтіні «Қазақ тілінің екі томдық түсіндірме сөздігі» бойынша алынған әріптердің жиілік сөздігіндегі ең жиі қолданылатын әріптердің алатын орны мен статистикасы жоғарыда қысқаша қарастырылған екі стиль бойынша алынған мәліметтерден белгілі дәрежеде айырым табатынын аңғаруға болады. Әсіресе, сөздіктегі тұйық раймен берілген етістіктердің әсерінен «у» әрпінің 3-орынға ие болуы және «қ» әрпінен басталатын қазақ сөздерінің сөздіктегі басымдылығы бірден көзге түседі. Ал





Т	6,19%	Г	1,20%
Р	5,62%	О	0,89%
Л	5,23%	Ө	0,89%
І	5,20%	Ә	0,88%
С	4,27%	У	0,83%
Д	4,27%	Ү	0,69%
М	3,42%	Я	0,37%
Қ	3,35%	В	0,26%
К	2,98%	Х	0,23%
О	2,44%	Ц	0,15%
Б	2,31%	Ф	0,12%
Ұ	2,18%	Ю	0,09%
Ж	1,81%	Э	0,08%
Ң	1,55%	Ь	0,07%
Ғ	1,54%	Ч	0,03%
З	1,47%	Һ	0,01%
		Е	0,00%

Ең жиі кездесетін әріп А – 12,52%, Е, Ы, Н, Т әріптері – шамамен 6 мен 7 %, Р, Л, І – 5%, С, Д, М, Қ әріптері – 3, 4 %, К, О, Б, Ү – 2%, қалған әріптер 2 %-ден төмен, яғни сирек немесе мүлде жоқтың қасы [6, 91 б.].

А.Шәріпбай аталған кітабында әріп жиілігін қазақ жазуын латын қарпіне көшірудегі емле ережелеріне қажетті статистикалық мәліметтер ретінде негізге алған. Автор бұл еңбегінде жаңа әліпбидің диграф нұсқасын ұсынып, әріп жиілігі бойынша У,Ө,Ә,Ү әріптерінің сирек кездесетінін тілге тиек етіп, оларды диграфтармен беру керек деген тұжырым жасайды. Сондай-ақ бұл әріптердің сирек кездесуі олардың тек түбір сөздерде ғана кездесетіндігі, яғни қосымшаларда мұндай дыбыстар болмауынан деп дұрыс тұжырым жасайды [6, 90 б.]. Ал І дыбысы жиі қолданылатындықтан, сөздің құрамы ұзын болып кетпес үшін, оны латынның Іі таңбасымен алу керек дегенді айтады. Автор қазақ дыбыстарына қатысты автоматты аударым жасауға қажетті емле ережелерін көрсеткен. Әрине автоматты аударма жасау үшін мұндай ереже керек, алайда автордың кейбір сөздердің дыбыстық құрамына қатысты емле ережелері тіл заңдылығына сәйкес келе бермейді [6, 94 б.].

Сонымен, жоғарыда қазақ тіліндегі әріп жиілігі бойынша жасалған зерттеулерге шолу жасалып, қазақ тіліндегі жиі кездесетін дыбыстар (әріптер) мен сирек кездесетін әріптердің жиілігі бойынша сипатталды. Зерттеулердегі жанрлық, екі тілдегі (қазақ және өзбек говорлары)

әріптердің жиілігі бойынша салыстырмалы талдаулардан әріп жиілігі бойынша негізінен ұқсас, барабар жиынтық шамалар алынғанын байқаймыз.

### 1.3 Қазақ әріптерінің жиілік сипаттамасы

Тіл – дамып отыратын динамикалық құбылыс. Сондықтан тілдегі кейбір құбылыстарды әркез қайта қарастырып, ондағы өзгерістер мен жаңалықтарды тіркеп, реттеп отыру қажет. Қазақ жазуын реформалау тек әліпби ауыстыру ғана емес, тілдік заңдылықтарды қайта жаңғырту мәселесімен де тығыз байланысты. Осы орайда қазақ жазуын латын қарпіне аударым жасауда тілдік бірліктер туралы эмперикалық сипаттағы статистикалық мәліметтер алу қазақ тілінің әліпби жүйесін ретке келтіріп, реформалауда аса қажет. Кезінде ЭЕМ-дердің жұмысы арқылы алынған статистикалық мәліметтер мен қазіргі миллиондаған сөзқолданыстан тұратын көлемді мәтіндерден алынған статистикалық мәліметтер ақпараттың шынайылығы мен дәлдігі жағынан айырым табатыны күмәнсіз. Бүгінгі таңда, осындай ақпараттық технология заманында ұлттық корпустар мен көлемді лингвистикалық базалармен жұмыс істеу күн тәртібіне қойылып отыр. Қазақ жазуын реформалау кезіндегі емле ережелерін жасау, орфографиялық сөздіктер құрастыру істерін жүзеге асыру үшін қазақ тіліндегі әріптен бастап сөзформаларына дейінгі тілдік бірліктердің жиілік сөздіктерін құрастырып, статистикалық зерттеулер жасау мақсат етіліп отыр.

Ұсынылып отырған қазақ әріптерінің жиілік сөздігі А.Байтұрсынұлы атындағы Тіл білімі институтының «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» мегажобасы бойынша орындалған «Жалпы білім берудегі қазақ тілінің жиілік сөздігінің» бір түрі ретінде алынған үлкен бір нәтиже – «Сөзформалардың жиілік сөздігінің» мәтінінен алынды [7.]. Бұл сөздікке оқыту курсының көбірек қамтитын әдеби тілдің көркем әдебиет, ресми-іскери, публицистикалық, сөйлеу сияқты бес түрлі стилі бойынша топтастырылған сан түрлі типтер мен жанрадағы әртүрлі авторға тән көлемі 7 миллион сөзқолданыстан тұратын ауқымдық база негіз болды [8,18.]. Демек, ұсынылып отырған қазақ әріптерінің жиілік сөздігіне алынған мәтін көлемі осы «Сөзформалар сөздігіндегі» 7 миллион сөзқолданыстан тұратын мәтін.

Қазақ әріптерінің жиілік сөздігі екі түрлі нұсқада жасалды. Бірі – «Қазақ тіліндегі әріптердің Әліпбилі жиілік сөздігі», екіншісі «Қазақ тіліндегі әріптердің жиілікті әліпбилі жиілік сөздігі». *7-кесте.*

7-кесте. «Қазақ тіліндегі әріптердің жиілікті әліпбилі жиілік сөздігі»  
(Статистикалық мәліметтер толық қамтылған)

№	Әріп	Абсолютті жиілік	Қосынды абсолютті жиілік	Қатынастық жиілік	Қосынды қатынастық жиілік	Бір топ әріптің мәтінді қамту пайызы
1	2	3	4	5	6	7
1	а	8630862	8630862	0,12796	0,12796	12,796
2	е	5660025	14290887	0,08391	0,21187	21,187
3	ы	5378832	19669719	0,07974	0,29161	29,161
4	н	4192570	23862289	0,06216	0,35377	35,377
5	і	4154893	28017182	0,06160	0,41537	41,537
6	т	4066134	32083316	0,06028	0,47565	47,565
7	р	4028709	36112025	0,05973	0,53537	53,537
8	л	3824239	39936264	0,05670	0,59207	59,207
9	д	3413312	43349576	0,05060	0,64267	64,267
10	с	2708334	46057910	0,04015	0,68283	68,283
11	м	2040371	48098281	0,03025	0,71308	71,308
12	қ	1876617	49974898	0,02782	0,74090	74,090
13	о	1639224	51614122	0,02430	0,76520	76,520
14	к	1633360	53247482	0,02422	0,78941	78,941
15	ғ	1256146	54503628	0,01862	0,80804	80,804
16	б	1229938	55733566	0,01823	0,82627	82,627
17	й	1175543	56909109	0,01743	0,84370	84,370
18	у	1103283	58012392	0,01636	0,86006	86,006
19	з	1095869	59108261	0,01625	0,87630	87,630
20	п	978676	60086937	0,01451	0,89081	89,081
21	ш	905698	60992635	0,01343	0,90424	90,424
22	г	903479	61896114	0,01339	0,91763	91,763
23	ң	872812	62768926	0,01294	0,93057	93,057
24	и	866761	63635687	0,01285	0,94342	94,342
25	ж	738252	64373939	0,01094	0,95437	95,437
26	ұ	673484	65047423	0,00998	0,96435	96,435
27	ө	655787	65703210	0,00972	0,97408	97,408



Жоғарыдағы кестеде әріптердің кездесу жиілігі абсолюттік жиілік пен бір топ әріптің мәтінді қамту пайызынан басқа, қосынды абсолюттік жиілік, қатынастық жиілік, қосынды қатынастық жиіліктері берілген.

Мұндай жиіліктерді мынадай формула арқылы көрсетуге болады.

**7 (А)-кесте. Қазақ тілі әріптерінің жиілікті әліпбилі сөздігі**

№ р/с (r) r=1, 2, 3, 4, 5,...	Әріп (әріп тіркесі)	Абсолютті жиілік (F <sub>r</sub> )	Қатынастық жиілік $f_r = \frac{F_r}{N}$	Қосынды Қатынастық жиілік $\sum_{r=1}^n f_r$	Бір топ әріптің мәтінді қамту пайызы $\sum_{r=1}^n f_r \times 100\%$
1	2	3	4	5	6
1	<i>A</i>	<i>F</i> <sub>1</sub>	<i>F</i> <sub>1</sub> / <i>N</i>	<i>F</i> <sub>1</sub> / <i>N</i>	( <i>F</i> <sub>1</sub> / <i>N</i> ) x100%
2	<i>E</i>	<i>F</i> <sub>2</sub>	<i>F</i> <sub>2</sub> / <i>N</i>	<i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i>	( <i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> ) x100%
3	<i>Ы</i>	<i>F</i> <sub>3</sub>	<i>F</i> <sub>3</sub> / <i>N</i>	<i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i>	( <i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i> ) x100%
4	<i>H</i>	<i>F</i> <sub>4</sub>	<i>F</i> <sub>4</sub> / <i>N</i>	<i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i> + <i>F</i> <sub>4</sub> / <i>N</i>	( <i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i> + <i>F</i> <sub>4</sub> / <i>N</i> ) x100%
5	<i>I</i>	<i>F</i> <sub>5</sub>	<i>F</i> <sub>5</sub> / <i>N</i>	<i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i> + <i>F</i> <sub>4</sub> / <i>N</i> + <i>F</i> <sub>5</sub> / <i>N</i>	( <i>F</i> <sub>1</sub> / <i>N</i> + <i>F</i> <sub>2</sub> / <i>N</i> + <i>F</i> <sub>3</sub> / <i>N</i> + <i>F</i> <sub>4</sub> / <i>N</i> + <i>F</i> <sub>5</sub> / <i>N</i> ) x100%
---	-----	-----	-----	-----	-----
		$N = \sum_{r=1}^n F_r$			<b>100%</b>

Кестедегі F<sub>r</sub> таңбасы әріптердің абсолютті жиіліктері, абсолютті жиіліктердің қосындысы мәтін көлемін N білдіреді. Қатынастық жиілік әрбір абсолюттік жиілікті осы қосынды жиілікке бөлінуі арқылы алынады. Ал қосынды қатынастық жиілікте қатынастық жиіліктер біртіндеп қосылып отырады. Ал әріптердің бір топ мәтінді қамту пайызын алу үшін 5-бағанадағы қосынды қатынастық жиілік 100%-ға көбейтіліп отырады.

Мұндай жиілік ақпараттар басқа алдағы сипатталатын жиілік сөздіктерде берілмеді.

Сонымен, бұның алдында жасалған әріптердің статистикалық талдауларындағыдай, қазақ тілінде ең жиі кездесетін әріп – **А** фонемасы. Одан кейінгі орында **Е, Ы** фонемалары. Ал **Ө, Ү, Ү** дыбыстарының сирек кездесуі қазақ тілінде ерін үндестігінің жазуда сақталмауынан болар деп ойлаймыз. Яғни бұл дыбыстар түбір сөздер құрамында болғанмен, қосымшалар құрамында езулік дыбыстармен жазылғандықтан, аз санды көрсетіп отыр. Ал дауыссыздар ішінде ең жиі кездесетін әріп – **Н** фонемасы, **Т, Р, Л** әріптері жиілігі бойынша екінші орында. 30-орында тұрған **Ә**-ден төмен тұрған әріптердің барлығы (**я, х, ц, ф, в, э, ь, ю, һ, ь, ч, щ, ё**) кірме әріптер. **Ә** әрпінің сирек кездесуі бір жағынан қосымшаларда кездеспеуінен болуы да мүмкін.

Жоғарыда айтылғандарды қорытындылай келе айтарымыз, қазақ тілінде ең жиі кездесетін **А, Е, Ы, І** дыбыстарын және сөз басында жиі кездесетін **Қ** дауыссыз дыбысын, сонымен қатар **Н, Т, Р, Д, Л, С** дауыссыз дыбыс әріптерін пернетақтада орналастыруда саусақпен басуда оңай басылып, икемді қимылдайтын саусақ саналатын сұқ саусақтың пернетақтадағы орналасуы мен басу ыңғайына қарай қойған дұрыс. Бұл пернетақтаның ортаңғы тұсына сәйкес келеді.

Сирек кездесетін жоғарыда аталған **В, Х, Ч, Һ, Щ, Ц** сияқты дауыссыз дыбыстар, **Э, Ю, Я** – дауыстыларын белгілейтін қаріптерді басуға икемсіз болып саналатын қол шетіндегі кішкентай шынашақ пен жанындағы аты жоқ саусақтың пернетақтадағы орналасу ретіне қарай, яғни пернетақтаның екі жақ шеткері бөлігіне орналастыру қажет.

Мұндай тәжірибе орыс пернетақтасын жасауда да негізге алынған. «Орыс пернетақтасында ең күшті әрі жылдам әрекет ететін сұқ саусақ басатын ауданға ең жиі кездесетін әріптер қатарын берген, ал әлсіз аты жоқ саусақ пен шынашақ басатын аудандарға сирек кездесетін әріптерді орналастырған» [9, 96.]. Ал түрік тәжірибесінде алдымен жасалған пернетақтада, керісінше, мәтінде жиі кездесетін әріптерді сұқ саусақпен басатын ауданнан қашығырақ орналастырған. Жиі кездесетін әріптерді сұқ саусақ басатын ауданнан қашық орналастырудың себебі саусаққа түсетін күштің жүктемесін азайту болған. Алайда түріктердің бұл пернетақтасы тиімсіз болып, кейін өзгертілген. Қазақ әліпбиін латын қарпіне көшіруде пернетақта жасау қажеттілігі туса, осындай әлемдік тәжірибелер мен қазақ әріптерінің қолданылу жиілігіне сүйеніп, тиімді пернетақта жасау қажет болады.

Жоғарыда сөз болған қазақ әріптерінің статистикалық мәліметтерді қазақтың ұлттық компьютерлік пернетақсына қазақ әліпбиін латынға көшіру кезінде пайдалануға болады.

## 1.4 Қазақ тіліндегі әріп тіркестерінің жиілік сөздігін құрастыру тарихы

Қазақ тіл білімінде әріп тіркестерін статистикалық тұрғыдан зерттеудің тарихы 1973 жылы жарық көрген «Қазақ тексінің статистикасы» атты еңбектен басталады деуге болады. Қ.Бектаев, А.Қ.Жұбановтың «Графема-фонема тіркестерінің жиілік тізімдері туралы», А.Қ.Жұбановтың «Қазақ текстеріндегі әріптер тіркестерінің статистикасы», Е.Агманов пен А.Қ.Жұбановтың «Распределения частот появления сочетаний знаков в орхоно-енисейской письменности», Қ.Бектаевтың «Қарақалпақ тіліндегі фонема тіркестерінің статистикасы» атты еңбектерін атап айтуға болады.

Әріп тіркесі бойынша Е.Агманов пен А.Жұбанов Орхон-Енисей ескерткіштерінің мәтіндегі әріптер тіркесімі бойынша жиілік статистикасын алуы көне түркі ескерткіштерін зерттеушілер үшін құнды ақпарат болып табылады [10, 630 б.].

Қ.Бектаев «Статистика фонемосочетаний в каракалпакском языке» атты мақаласында қарақалпақ тіліндегі фонема тіркесімінің статистикасын жасауда екі әріпті тіркесімнің жалпы жиілігін және сонымен қатар сөз басы мен сөз соңында кездесетін позициялық жиілігін де анықтайды [11, 365.]. Мәтін 13921 сөзқолданыстан тұратын Н.А.Баскаковтың «Қарақалпақский язык» (1951) атты кітабынан алынады.

Қ.Бектаев, А.Қ.Жұбановтың «О частотных списках графемно-фонемных сочетаний» деп аталатын мақалада жоғарыдағы Орхон-Енисей ескерткіштеріндегі әріп тіркесінің жиілігін алуға арналған мәтін Күлтегін ескерткішінен алынғанын (сөзқолданыс саны – 1635, таңбақолданыс саны – 8840), сондай-ақ әріп тіркестерінің жиілігі сөз басы мен сөз соңына қатысты да жасалғандығын айтады.

А.Қ.Жұбановтың «Қазақ текстеріндегі әріптер тіркестерінің статистикасы» деп аталатын жиілік сөздікте жоғарыда әріп жиіліктерін үш жанр бойынша салыстырмалы статистика жасалған. Алынған мәтіндер мен көлемі: «Абай жолы» романынан – 51290 сөзқолданыс, 280843 әріпқолданыс, қазақ тілінің екітомдық түсіндірме сөздігінен – 17585 сөзқолданыс, 116317 әріпқолданыс, ғылыми-техникалық әдебиеттен – 19467 сөзқолданыс, 130188 әріпқолданыс. Мұнда әріп тіркестерінің жиілігімен қатар, ең жиі кездесетін әріп тіркестерінің сөздің басқа және соңғы позицияларындағы жиілігі де берілген.

Осы жиілік сөздікте жиілігі 3500-ден асатын әріп тіркестерінің тізімін көрсетсек [12, 614 б.], **8-кесте**.





7	ДЕ	33340
8	ДА	33279
9	АЛ	33022
10	КЕ	30610

Көріп отырғанымыздай, 8 және 9-кестелердегі жиі кездесетін әріп тіркестері бірлі-жарым болмаса, негізінен ұқсас.

### 1.5 Әріп тіркестерінің жиілік сипаттамасы

Ұсынылып отырған қазақ әріптер тіркесінің жиілік сөздігі А.Байтұрсынұлы атындағы Тіл білімі институтының «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» мегажобасы бойынша орындалған «Жалпы білім берудегі қазақ тілінің жиілік сөздігінің» бір түрі ретінде алынған үлкен бір нәтиже – «Сөзформалардың жиілік сөздігінің» мәтінінен алынды [7.]. Қазақ әріп тіркестерінің жиілік сөздігіне алынған мәтін көлемі осы «Сөзформалар сөздігіндегі» 7 миллион сөзқолданыстан тұратын мәтін.

Қазақ әріп тіркестерінің жиілік сөздігі екі түрлі нұсқада жасалды. Бірі – «Қазақ тіліндегі әріп тіркестерінің Әліпбилі жиілік сөздігі», екіншісі – «Қазақ тіліндегі әріп тіркестерінің жиілікті әліпбилі жиілік сөздігі».

«Қазақ тіліндегі әріп тіркестерінің Әліпбилі жиілік сөздігінде» әріп тіркестері әліпби ретімен берілген. Ал «Қазақ тіліндегі әріп тіркестерінің жиілікті әліпбилі жиілік сөздігінде» жиі кездесетін тіркестер алдымен беріліп, кему жиілігі бойынша тізімделеді.

Енді қазақ әріптерінің екіәріптік тіркестерінің статистикалық сипаттама-сына қысқаша тоқталайық. **10-кесте.**

**10-кесте.** «Қазақ тілінің екіәріптік тіркестерінің жиілік сөздігінен» үзінді

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
1	<i>ар</i>	688045	2,035495
2	<i>ан</i>	582201	3,757865
3	<i>ал</i>	565089	5,429610
4	<i>ын</i>	562671	7,094202
5	<i>да</i>	560939	8,753670
6	<i>та</i>	511161	10,265877
7	<i>ен</i>	501533	11,749600

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
8	<i>ер</i>	481450	13,173909
9	<i>қа</i>	460416	14,535993
10	<i>де</i>	458037	15,891038
11	<i>ін</i>	451434	17,226550
12	<i>ла</i>	422562	18,476647
13	<i>ды</i>	417831	19,712748
14	<i>ты</i>	404750	20,910150
15	<i>нд</i>	399330	22,091518
16	<i>га</i>	396123	23,263399
17	<i>ке</i>	371958	24,363790
18	<i>ай</i>	369359	25,456493
19	<i>лы</i>	342672	26,470245
20	<i>ті</i>	333046	27,455520
21	<i>ді</i>	331481	28,436166
22	<i>ас</i>	330440	29,413731
23	<i>сы</i>	327308	30,382031
24	<i>ме</i>	323281	31,338418
25	<i>ат</i>	284740	32,180786
26	<i>ба</i>	281240	33,012799
27	<i>ны</i>	276945	33,832107
28	<i>ың</i>	275833	34,648124
29	<i>ол</i>	273816	35,458175
30	<i>те</i>	273269	36,266607
31	<i>ет</i>	272744	37,073486
32	<i>ры</i>	263800	37,853906
33	<i>ық</i>	261904	38,628716
34	<i>ра</i>	257203	39,389619
35	<i>на</i>	250705	40,131299
36	<i>жа</i>	248531	40,866547
37	<i>ақ</i>	243641	41,587328
38	<i>ыл</i>	241848	42,302805
39	<i>ле</i>	241186	43,016324
40	<i>не</i>	239816	43,725790
41	<i>ст</i>	239516	44,434368

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
42	<i>ір</i>	237178	45,136030
43	<i>рі</i>	230151	45,816903
44	<i>ма</i>	229464	46,495744
45	<i>са</i>	226934	47,167100
46	<i>ге</i>	226772	47,837976
47	<i>ел</i>	209952	48,459093
48	<i>ек</i>	208945	49,077231
49	<i>ғы</i>	207327	49,690582
50	<i>лд</i>	198170	50,276843
51	<i>ыр</i>	197109	50,859966
52	<i>ағ</i>	196589	51,441550
53	<i>рд</i>	195039	52,018548
54	<i>ес</i>	189725	52,579826
55	<i>іл</i>	189148	53,139397
56	<i>тт</i>	184683	53,685759
57	<i>бі</i>	184178	54,230626
58	<i>ің</i>	181594	54,767850
59	<i>лі</i>	181226	55,303984
60	<i>гі</i>	178293	55,831442
61	<i>ау</i>	177800	56,357441
62	<i>қы</i>	172081	56,866521
63	<i>ре</i>	169334	57,367475
64	<i>ыс</i>	168863	57,867035
65	<i>сі</i>	168801	58,366412
66	<i>ып</i>	166290	58,858360
67	<i>ні</i>	163289	59,341430
68	<i>ам</i>	161755	59,819962
69	<i>ік</i>	154050	60,275700
70	<i>ым</i>	147729	60,712738
71	<i>ім</i>	147386	61,148761
72	<i>бо</i>	146149	61,581124
73	<i>қт</i>	144973	62,010009
74	<i>ад</i>	144878	62,438613
75	<i>бе</i>	143874	62,864246

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
76	<i>шы</i>	140779	63,280723
77	<i>се</i>	137879	63,688621
78	<i>ем</i>	133338	64,083085
79	<i>із</i>	131926	64,473371
80	<i>кі</i>	130520	64,859499
81	<i>ег</i>	128229	65,238848
82	<i>кө</i>	127963	65,617411
83	<i>ор</i>	126556	65,991811
84	<i>ап</i>	120653	66,348748
85	<i>мы</i>	120124	66,704120
86	<i>ыз</i>	119848	67,058675
87	<i>ұр</i>	118474	67,409166
88	<i>ша</i>	113697	67,745524
89	<i>ей</i>	113265	68,080604
90	<i>зд</i>	113149	68,415342
91	<i>аз</i>	113080	68,749875
92	<i>өз</i>	112899	69,083872
93	<i>қо</i>	112540	69,416808
94	<i>рт</i>	107298	69,734236
95	<i>ін</i>	105436	70,046155
96	<i>ші</i>	105243	70,357503
97	<i>он</i>	104735	70,667349
98	<i>рл</i>	104106	70,975334
99	<i>же</i>	103331	71,281025
100	<i>ед</i>	102174	71,583294
101	<i>йы</i>	98085	71,873467
102	<i>лг</i>	96179	72,158000
103	<i>кт</i>	95018	72,439099
104	<i>па</i>	92954	72,714092
105	<i>за</i>	91058	72,983476
106	<i>мі</i>	90191	73,250295
107	<i>ыг</i>	89655	73,515528
108	<i>зі</i>	89577	73,780530
109	<i>уы</i>	89453	74,045166

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
110	<i>мд</i>	87990	74,305473
111	<i>іс</i>	86953	74,562713
112	<i>со</i>	86002	74,817139
113	<i>йт</i>	85321	75,069551
114	<i>йд</i>	84808	75,320445
115	<i>ең</i>	83836	75,568463
116	<i>ту</i>	82158	75,811517
117	<i>тү</i>	81604	76,052932
118	<i>ой</i>	81473	76,293960
119	<i>ру</i>	78168	76,525210
120	<i>ұл</i>	78080	76,756200
121	<i>құ</i>	77632	76,985865
122	<i>рм</i>	77476	77,215068
123	<i>еп</i>	77425	77,444120
124	<i>ия</i>	76718	77,671081
125	<i>ше</i>	75995	77,895903
126	<i>аң</i>	69420	78,101273
127	<i>ос</i>	68825	78,304883
128	<i>ән</i>	67263	78,503873
129	<i>үр</i>	67191	78,702649
130	<i>рг</i>	66537	78,899490
131	<i>нт</i>	64861	79,091373
132	<i>рг</i>	64652	79,282638
133	<i>өр</i>	63611	79,470824
134	<i>оқ</i>	62543	79,655849
135	<i>аб</i>	62542	79,840872
136	<i>то</i>	61576	80,023037
137	<i>әр</i>	61565	80,205169
138	<i>бұ</i>	60962	80,385518
139	<i>йл</i>	60354	80,564068
140	<i>уд</i>	60316	80,742505
141	<i>жы</i>	59382	80,918179
142	<i>кү</i>	59169	81,093223
143	<i>жо</i>	58694	81,266862



Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
178	<i>су</i>	40052	86,144812
179	<i>үн</i>	39988	86,263112
180	<i>би</i>	39980	86,381388
181	<i>ри</i>	39419	86,498004
182	<i>еш</i>	39268	86,614173
183	<i>пе</i>	38952	86,729408
184	<i>уш</i>	38833	86,844290
185	<i>ие</i>	38602	86,958490
186	<i>от</i>	38302	87,071801
187	<i>мә</i>	37581	87,182980
188	<i>ко</i>	37456	87,293789
189	<i>зе</i>	37193	87,403820
190	<i>тқ</i>	37160	87,513753
191	<i>си</i>	36662	87,622213
192	<i>ыт</i>	35854	87,728283
193	<i>уг</i>	35743	87,834024
194	<i>ал</i>	35728	87,939721
195	<i>ми</i>	35603	88,045048
196	<i>ул</i>	35192	88,149159
197	<i>сп</i>	34919	88,252462
198	<i>ом</i>	33837	88,352565
199	<i>рн</i>	33105	88,450502
200	<i>жұ</i>	33023	88,548196
201	<i>ңі</i>	32705	88,644950
202	<i>оғ</i>	32191	88,740183
203	<i>ял</i>	32129	88,835233
204	<i>нт</i>	31942	88,929729
205	<i>ұм</i>	31057	89,021608
206	<i>ци</i>	30876	89,112950
207	<i>уі</i>	30113	89,202036
208	<i>іш</i>	29962	89,290675
209	<i>өн</i>	29755	89,378701
210	<i>ли</i>	29713	89,466603
211	<i>үл</i>	29290	89,553254

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
212	<i>ск</i>	29049	89,639192
213	<i>мұ</i>	29004	89,724997
214	<i>ая</i>	28939	89,810609
215	<i>лс</i>	28186	89,893994
216	<i>нг</i>	27874	89,976456
217	<i>ңы</i>	27579	90,058045

Аталған екіәріптік жиілік сөздікте әртүрлі тіркестердің рет саны 1160-қа тең, ал осы екіәріптік тіркестер қайталанып қолданыла келіп, олардың абсолютті жиіліктерінің қосындысы 67 604670 мәтін бойындағы екіәріптік тіркестердің 100 пайызын құрайды.

Енді осы сөздіктің жоғары жиілікті бөлігінен үзінді ретінде барлық мәтіннің 90 пайызын ғана қамтитын жоғары жиілікті зонаға жататын 217 екіәріптік тіркестер бөлігін **10-кесте** бойынша қарастыруға болады.

Қазақ тілінің ұлттық компьютерлік пернетақтасында **7-кестедегі** жиі кездесетін 1-топтағы 19 әріп екі бөлікке бөлініп, пернетақтаның ортаңғы бөлігіне орналасады. Сол сияқты, **10-кестедегі** екіәріптік тіркестердің ең жиі қолданылатын, мәселен, мәтіннің 90 пайызын қамтитын рет сандары 1-217 аралығын қамтитын тіркестердің екінші әріптерін **7-кестедегі** рет сандары 1-29 аралығындағы әріптерге тіркесуге пернетақтаның жақын аралығына, яғни сол қол мен оң қол саусақтарына сәйкестендіріп орналастыру мүмкіндіктері қарастырылды. Әрине, бұл әрекетке үлкен мән берілуі қажет деп білеміз.

Статистикалық мәліметтер толық болу үшін біз өз зерттеуімізде сөз басында және сөз соңында кездесетін екіәріптік тіркестерінің де жиілік сөздіктерін құрастырдық. Енді ол сөздіктердің үзінділерімен **11-кесте** және **12-кесте** арқылы танысуға болады.

**11-кесте.** Қазақ тілінің сөз басындағы екіәріптік тіркестерінің жиілік сөздігі (үзінді)

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
1	<i>қа</i>	314514	5,33657
2	<i>жа</i>	228241	9,20929
3	<i>ба</i>	222845	12,99045
4	<i>ке</i>	177014	15,99396
5	<i>бі</i>	149781	18,53540



Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
6	<i>бо</i>	144030	20,97925
7	<i>та</i>	142349	23,39458
8	<i>кө</i>	126340	25,53828
9	<i>ме</i>	123027	27,62576
10	<i>са</i>	116450	29,60164
11	<i>де</i>	116241	31,57398
12	<i>қо</i>	108225	33,41031
13	<i>ал</i>	99351	35,09607
14	<i>бе</i>	91785	36,65344
15	<i>со</i>	81582	38,03770
16	<i>же</i>	75181	39,31335
17	<i>құ</i>	73538	40,56111
18	<i>тү</i>	70582	41,75873
19	<i>ай</i>	65318	42,86702
20	<i>бұ</i>	59954	43,88430
21	<i>ор</i>	58427	44,87567
22	<i>қы</i>	58085	45,86124
23	<i>кү</i>	57775	46,84154
24	<i>жо</i>	57314	47,81403
25	<i>тұ</i>	55624	48,75784
26	<i>ма</i>	55004	49,69113
27	<i>ар</i>	54779	50,62060
28	<i>жә</i>	53475	51,52795
29	<i>шы</i>	51825	52,40730
30	<i>өз</i>	50541	53,26486
31	<i>те</i>	50297	54,11828
32	<i>жү</i>	49984	54,96639
33	<i>ат</i>	49131	55,80003
34	<i>ті</i>	48029	56,61497
35	<i>жы</i>	47056	57,41340
36	<i>ек</i>	45959	58,19322
37	<i>сө</i>	43937	58,93873
38	<i>ша</i>	43897	59,68356
39	<i>се</i>	42067	60,39734



Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
74	<i>ар</i>	22221	77,88167
75	<i>ел</i>	21756	78,25082
76	<i>ұл</i>	20966	78,60657
77	<i>ал</i>	20143	78,94835
78	<i>іс</i>	19666	79,28203
79	<i>бө</i>	19398	79,61117
80	<i>ем</i>	18988	79,93335
81	<i>ед</i>	18731	80,25118
82	<i>мү</i>	18613	80,56700
83	<i>сұ</i>	18011	80,87260
84	<i>дә</i>	17976	81,17761
85	<i>үй</i>	17527	81,47500
86	<i>ан</i>	17500	81,77194
87	<i>би</i>	17496	82,06880
88	<i>на</i>	17402	82,36407
89	<i>ен</i>	17090	82,65405
90	<i>әд</i>	16843	82,93984
91	<i>ты</i>	16671	83,22271
92	<i>су</i>	16160	83,49690
93	<i>өл</i>	15674	83,76286
94	<i>өн</i>	15137	84,01970
95	<i>си</i>	15010	84,27438
96	<i>іш</i>	14993	84,52878
97	<i>ұл</i>	14866	84,78102
98	<i>ұс</i>	14847	85,03294
99	<i>пр</i>	14819	85,28438
100	<i>кә</i>	14494	85,53031
101	<i>әк</i>	13795	85,76438
102	<i>ти</i>	13599	85,99512
103	<i>өм</i>	13112	86,21760
104	<i>Сә</i>	12973	86,43772
105	<i>бү</i>	12953	86,65751
106	<i>бә</i>	12880	86,87605
107	<i>жи</i>	12741	87,09224



**12-кесте. Қазақ тілінің сөз соңындағы екіәріптік тіркестерінің жиілік сөздігі (үзінді)**

<b>Рет саны №</b>	<b>Әріп тіркесі</b>	<b>Абсолютті жиілік</b>	<b>Бір топ әріп тіркестерінің мәтінді қамту пайызы</b>
1	<i>ен</i>	319537	5,41253
2	<i>ан</i>	272130	10,02205
3	<i>ың</i>	240700	14,09919
4	<i>ды</i>	240100	18,16617
5	<i>ын</i>	195229	21,47309
6	<i>ді</i>	187867	24,65530
7	<i>ін</i>	176123	27,63859
8	<i>ық</i>	165604	30,44370
9	<i>ің</i>	157597	33,11319
10	<i>ар</i>	149557	35,64648
11	<i>ып</i>	146130	38,12173
12	<i>да</i>	130213	40,32737
13	<i>ай</i>	113729	42,25378
14	<i>не</i>	109065	44,10120
15	<i>де</i>	102506	45,83752
16	<i>ға</i>	101192	47,55157
17	<i>ты</i>	98951	49,22767
18	<i>ік</i>	95044	50,83759
19	<i>ер</i>	91362	52,38514
20	<i>на</i>	89806	53,90634
21	<i>сы</i>	89508	55,42248
22	<i>ғы</i>	80934	56,79340
23	<i>ті</i>	80568	58,15811
24	<i>ге</i>	80523	59,52207
25	<i>ін</i>	78056	60,84423
26	<i>ақ</i>	75244	62,11876
27	<i>гі</i>	69045	63,28829
28	<i>лы</i>	68298	64,44517
29	<i>рі</i>	64818	65,54310
30	<i>ір</i>	64774	66,64029
31	<i>ап</i>	64127	67,72651
32	<i>еп</i>	56991	68,69187

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
33	<i>ыз</i>	53348	69,59551
34	<i>ры</i>	51962	70,47568
35	<i>із</i>	47858	71,28633
36	<i>ны</i>	47161	72,08517
37	<i>ек</i>	46525	72,87324
38	<i>ім</i>	44391	73,62517
39	<i>ау</i>	43245	74,35768
40	<i>қа</i>	43056	75,08699
41	<i>ыс</i>	42655	75,80951
42	<i>ым</i>	39675	76,48155
43	<i>ша</i>	37192	77,11154
44	<i>ет</i>	36345	77,72717
45	<i>лі</i>	35812	78,33378
46	<i>та</i>	35423	78,93380
47	<i>ұл</i>	32593	79,48588
48	<i>сі</i>	32030	80,02843
49	<i>ру</i>	31677	80,56499
50	<i>ес</i>	30999	81,09007
51	<i>ні</i>	30949	81,61431
52	<i>ке</i>	30465	82,13035
53	<i>са</i>	29680	82,63309
54	<i>ол</i>	27248	83,09463
55	<i>ші</i>	26643	83,54593
56	<i>ей</i>	26551	83,99567
57	<i>се</i>	25855	84,43362
58	<i>ыр</i>	24862	84,85474
59	<i>ат</i>	24729	85,27362
60	<i>ра</i>	24671	85,69152
61	<i>ас</i>	23516	86,08985
62	<i>ам</i>	23501	86,48792
63	<i>те</i>	23009	86,87766
64	<i>ла</i>	22852	87,26475
65	<i>ма</i>	22668	87,64871
66	<i>шы</i>	22100	88,02306

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
67	<i>кі</i>	22094	88,39730
68	<i>ту</i>	21449	88,76062
69	<i>уы</i>	20591	89,10940
70	<i>ше</i>	19712	89,44330
71	<i>ыл</i>	18768	89,76120
72	<i>оқ</i>	17887	90,06418

Енді оларды қазақ тілінің компьютерлік ұлттық пернетақтасын жасауда қалай пайдаланамыз деген сұрақ туындауы мүмкін. Осы сауалға қысқаша жауап берсек, біздің ұстанымыз бойынша ең жиі қолданылатын қазақ әріптері пернетақтаның ортаңғы жағында орналасуы қажет. Олар екі топқа бөлініп, бірінші, сол жақ қол саусақтарына және екінші, оң жақ қол саусақтарына деп ажыратылуы қажет. Сонымен бірге, екіәріптік тіркестердің статистикасы бойынша тіркестің екінші әрпі, тіркестің бірінші әрпінің оң жағына жақын орналасуы қажет деп білеміз. Ондай дәрежеге бір емес, бірнеше әріп «үміткер» болуы мүмкін. Осындай жағдайда екіәріптік тіркестің позициялық жағдайына да мән берілуі керек.

### 1.6 Ұлттық қазақ әліпбиін пернетақтада орналастырудың статистикалық негіздері

Әліпби ауыстыруда маңызды мәселелердің бірі – әріптерді пернетақтада орналастыру. Адамның әріптерді пернетақтаның қай бұрышына орналастырса да жаттанды түрде үйреніп алуы заңды. Алайда пернетақтада мәтін теруде әріптерді оңтайлы орналастыру адамның күш-қуатын және теруге кететін уақытты үнемдейді. Сондықтан ұлттық әліпбидегі әріптерді пернетақтада тиімді орналастыруда лингвистикалық, әлеуметтік зерттеулер жүргізу қажет. Лингвистикалық зерттеулер жүргізу негізінен тіл білімінің қолданбалы бағытының зерттеу нысаны. Бұл ретте әріптерді пернетақтада орналастыруда қандай ұстанымдарға сүйенеміз деген мәселенің басын ашып алу қажет.

Оның *біріншісі* – пернетақтадағы ағылшынша жазатын латын қаріптерінің тұрған орнын сақтап, кирилдегі қазақша әріптер тұрған жоғарғы жақ қатарға жаңа таңбаларды орналастыру. Бұл ағылшын пернетақтасында тұрған латын қаріптерін басуда шатастырмау үшін тиімді деуге болады. Кирилл қаріптері тұрған пернетақтада да қазақ тіліне ғана тән әріптер осылай жоғарғы қатарға қатар-қатар орналастырылған болатын.

*Екіншісі* – ағылшын пернетақтасындағы латын қаріптерін сақтай отырып, жоғары жаққа емес, негізгі үш қатардағы бос орындарға жаңа таңбаларды қою. Осы ұстанымға сүйеніліп, төмендегідей пернетақта жасалды.

**1-сурет. Ағылшындағы бос орындарға қойылған пернетақта**

Esc	~	!	"	£	\$	%	^	&	'	(	)	0	-	=	Bksp	Home	PgUp
Tab	q	w	e	r	t	y	u	i	o	p	á	ñ	~	#	Del	End	PgDn
Caps	a	s	d	f	g	h	j	k	l	g	ú				Insert	Pause	
Shift		\	z	x	c	v	b	n	m	ú	ó			'	Shift	PrtScn	ScLk
Ctrl	#	Alt							Alt	®	Ctrl	~	~	~	Fn	Параметры	Справка

Бұл пернетақтада ағылшындағы бос орындарға жаңаша таңбаланған ә, ң, ғ, у, ү, ө (á, ñ, g, ú, ó) әріптері қойылды. Олар пернетақтаның оң жақ қол бетіндегі ортаңғы, жоғарғы және төменгі қатарларға екі-екіден орналастырылды. Бұл пернетақтада әзірге И/Й (I) әрпі орналастыртырылмады. Себебі, соңғы әліпбиде I мен И/Й әріптері бірдей дерлік болып тұр.

*Үшіншісі* – әлемдік тіл білімінде ұстанылатын принцип – пернетақтада әріптерді қолданыста жиі кездесуіне қарай орналастыру, яғни жиі кездесетін әріптерді пернетақтада басқанда оңай, икемді басылатын жерлерге қою болып табылады.

Әріптерді пернетақтада тиімді орналастыру мәселесіне қатысты А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімі қолданыстағы әріптердің кездесу жиілігін анықтады. Әріп жиілігі туралы мәліметтер (9-кесте) жоғарыдағы тақырыптарда қарастырылған.

Әріптерді пернетақтада орналастыруда әріп жиілігінен басқа, олардың қасындағы келесі әріппен тіркесуін де негізге алу пернетақта батырмасын басудың тиімділігі үшін қажет болғандықтан, олардың оң жақтарында жоғарыда үзінді келтірілген (10, 11, 12-кестелер) екіәріптік тіркестер жиілік сөздіктерінен алынған деректер мәліметіне сүйенген жөн болады.

Қазақ тілінің ұлттық компьютерлік пернетақтасында жиі кездесетін **19** әріп екі бөлікке бөлініп, пернетақтаның ортаңғы бөлігіне орналасады. Сол сияқты **10-кестедегі** екіәріптік тіркестердің ең жиі қолданылатын тіркестердің екінші әріптерін жиі кездесетін әріптерге тіркесуге пернетақтаның жақын аралығына, яғни сол қол мен оң қол саусақтарына сәйкестендіріп орналастыру қажет.

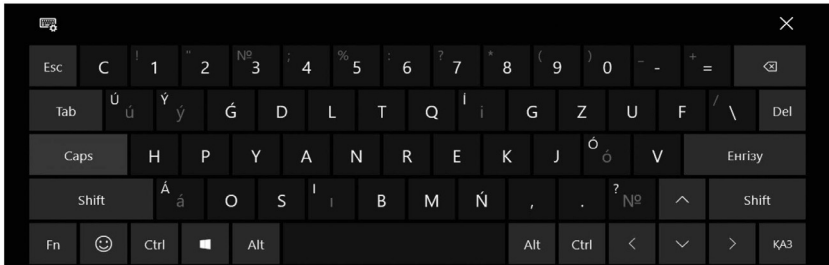
Осы жиілік талдау (әріп және әріп тіркесінің жиілік сөздіктері) нәтижесі бойынша қазақ әріптері мен әріп тіркестерінің жиі кездесуі ескеріліп орналастырылған пернетақтаның пилоттық жобасы ұсынылды.



Пернетақтада әрбір әріптің кирилдегі және латындағы баламалары келтірілді. Пернетақтаның ортаңғы бөлігінде жиілік сөздіктегі ең жиі қолданылған әріптер орын алды, ал олардың оң жақтарында екіәріптік тіркес жиілік сөздігінен алынған әріптер орналастырылды.

Пернетақтаның шеткі жақтарына сирек кездескен әріптер орналастырылды. Пернетақтада көрініс таппаған дыбыстар **ш, ч** мынадай әріп тіркестерінің **sh/ш, ch/ч** әрбіреуінің пернетақтада бар болуынан, олар орын алмады.

### 2-сурет. Статистикалық негіздегі пернетақта



Енді қолданыстағы әріптің жиілігі мен олардың тіркесіп келу жиілігі ескерілген осы пернетақтаға орналастырылған әрбір әріптің не үшін осылай орналастырылғанын баяндаймыз.

Пернетақта батырмалары екі қолмен басылатындықтан, ортасынан оңға және солға қарай есептеліп, екі қолға бөлінеді. Пернетақтаның шекарасы қазіргі кирилл пернетақтасындағы А және О әріптері тұрған кертікпен есептеледі. Осы екі әріптің ортасында тұрған П және Р әріптерінің ортасы орталық (центр) болып саналады. Яғни ЕПИ әріптерінің оң жағы, НРТ әріптерінің сол жағы.

### 3-сурет. Пернетақтаның шекара кертпесі мен ортаңғы бөлігі



Әріптерді пернетақтада орналастыруда оның бетіндегі батырмалар орналастырылған қатардың басуға, яғни мәтін теруге ыңғайлылығына қарай актив және пассив аймақтарға бөлінеді. Пернетақтаның әріп теруге ыңғайлы жері ортаңғы аймақ, әсіресе керткі тұрған жердің айналасы, асты мен үсті болып табылады. Ал шеткі батырмалар пассив аймаққа жатады. Актив және пассив аймақтарды пернетақтаның горизонталь (көлденең) және вертикаль (тіке) бөліктері бойынша да көрсетуге болады. Көлденең жеке бойынша актив аймақ – ол екі керткі тұрған ортаңғы қатар, яғни қазіргі кирилдегі Ф әрпінен басталып, Э әрпіне аяқталатын қатар. Одан кейінгі актив аймақ осы қатардың үстінде орналасқан үстіңгі қатар. Ал астыңғы қатар көбінесе сиректеу кездесетін әріптер орналасатын аймақ.

Батырмалардың тік бағыттағы (вертикальды) межесі керткі тұрған батырмалардың жаны, яғни ортаңғы, екінші қатардың жаны, ортаңғы тұсы. Сонымен қатар үстіңгі қатардағы кертпенің үстіңгі ортаңғы бөліктері. Одан кейінгісі кертпенің астыңғы ортаңғы бөліктері.

Ал осы үш қатардың шеткі бөліктері пассив аймақ болып есептеліп, сирек кездесетін әріптер не тыныс белгілер орналастырылады.

Осы шарттарға сәйкес біз әрбір әріптің кездесу жиілігі мен басқа әріптермен тіркесу жиілігіне қарай пернетақтаға орналастырдық.

**А (А)** әрпі – жиілік сөздік бойынша, 1-орында тұрған ең жиі кездесетін әріп. Біздің пернетақтада А әрпі басуға оңтайлы болып саналатын ортаңғы қатарға бұрынғы кирилдегі А әрпінің орнына қойылды.

Жиілік сөздікте жиілігі жағынан 2-орында тұрған **Е (Е)** әрпі осыған сәйкес оң қолға екінші кертпенің үстіне қойылып отыр.

Жиілік сөздікте 3-орында тұрған **Ы (Y)** әрпі де ортаңғы қатарға А-ның сол жағына орналастырылды.

**Н (N)** әрпі жиілігі бойынша 4-орында тұр. Сонымен қатар әріп тіркесінің жиілік сөздігінде А-дан кейін жиі тіркесетін әріп (АН). Сондықтан N әрпін А әрпінен кейін қою статистикалық тұрғыдан дұрыс болып табылады.

**І (I)** әрпі – жиілігі бойынша 5-орында. Сондықтан ортаңғы бөліктің жоғарғы қатарына Е-нің үстіне, оң жақ қолға латынның бұрынғы І-сі тұрған батырмаға орналастырылды.

**Т (T)** әрпі – жиілігі бойынша 6-орында, әріп тіркесі бойынша да тіркес басында, 6-орында (та) тұр. Сондықтан үстіңгі орталық бөлікке орналастырылды.

**Р (R)** әрпі – әріп жиілігі бойынша 7-орында, ал әріп тіркесінде алдыңғы позицияда (ар) 1-орында тұр. Сондықтан Е-нің жанында ортаңғы бөлікке, кертпенің сол жағына, яғни орталыққа орналастырылды.

**Л (L)** әрпі – әріп жиілігі бойынша 8-орында тұр. Алайда әріп тіркесі бойынша 3-орында. Сондықтан А-ның дәл үстінде орталық бөлікте, үстіңгі қатарда тұр.

**Д (D)** әрпі – әріп жиілігі бойынша 9-орында, ал әріп тіркесі бойынша 5-орында, алдыңғы позицияда (ДА). Сондықтан бұл әріп те жиі кездесетін әріптер қатарынан орын алып, орталық бөлікте кертпеде тұрған А-ның үстіне, сол жақ қолға, жоғарғы қатарға беріліп отыр. Бұл жерде Д-ның А әрпімен тіркесу жиілігі де ескеріліп, жақын жерден орын берілген.

**С (S)** әрпі – әріп жиілігі бойынша 10-орында, әріп тіркесінде соңғы позицияда 22-орында тұр (АС). Жиілігі бойынша белсенділер қатарына жағқанымен, тіркесімі бойынша аса белсенді емес болғандықтан, астыңғы қатарға А-ның астыңғы жағына, бұрынғы латын мен кирилдегі орынға қойылды.

**М (M)** әрпі – әріп жиілігі бойынша 11-орында, әріп тіркесінде 24-орында бастапқы позицияда (МЕ). Жиі кездесетіндер қатарына жататындықтан, ортаңғы бөлікте, үтір мен нүкте батырмалары қойылған батырманың қасында, шетке қарай оң жақ қолға астыңғы қатарда орналастырылды.

**Қ (Q)** әрпі – әріп жиілігі бойынша 12-орында, әріп тіркесінде бастапқы позицияда (ҚА) 9-орында. Сондықтан жиі басылатын әріптер жанында үстіңгі қатарда орналастырылды.

**О (O)** әрпі – әріп жиілігінде 13-орында, әріп тіркесі бойынша 29-орында, яғни өте жиі кездеспейтіндіктен, астыңғы қатарға, шеткі сол жақ бөлікке қойылды.

**К (K)** әрпі – әріп жиілігі бойынша 14-орында, әріп тіркесі бойынша 17-орында, бастапқы позицияда (КЕ). Тіркесте Е әрпімен жиі кездесетін болғандықтан, кертінде тұрған Е-нің жанына сол жақ қолға орналастырылды.

**Ғ (G)** әрпі – әріп жиілігі бойынша 15-орында, әріп тіркесінде 16-орында бастапқы позицияда тұр. Өте жиі кездесетін әріптерге жатпайтындықтан, шеткі жоғарғы қатарға орналастырылды.

**Б (B)** әрпі – әріп жиілігі бойынша 16-орында, тіркесте 26-орында, бастапқы позицияда тұр. Сондықтан орталық бөлік болса да астыңғы қатарға, бұрынғы латындағы В-ның орнына қойылды.

**Й мен И** әрпінің жиілігі қосылады. Өйткені жаңа әліпбиде бұл екеуі бір әріппен таңбаланады. Й әрпі – әріп жиілігі бойынша 17-орында, әріп тіркесінде 18-орында, соңғы позицияда (АЙ). Екі әріп жиілігі қосылатындықтан, жиі кездесетін әріп ретінде есептеліп, пернетақтаның ортаңғы бөлігіне лайық, алайда астыңғы қатарда, сонымен бірге А-мен кездесуі жиі болғандықтан, А-ның астында тұр.

**Ү (Y)** әрпі – әріп жиілігінде 18-орында, әріп тіркесінде 61-орында, соңғы (екінші) позицияда (АҮ). Яғни сирек кездеседі. Сондықтан шеткі сол жаққа үстіңгі қатарға берілді.

**З (Z)** әрпі – әріп жиілігінде 19-орында, әріп тіркесінде 79-орында екінші позицияда (ІЗ). Бұл әріп сирек кездесетін әріпке жатады. Сондықтан үстіңгі шеткі орынға қойылды.

**П (P)** әрпі – әріп жиілігі бойынша 20-орында, тіркесі бойынша 66-орында, екінші позицияда (ЫП) тұр. Бұл да сирекке жатады. Сондықтан шетке шығарылды, алайда ортаңғы қатарда.

**Ш (SH)** әрпі – әріп жиілігі бойынша 21-орында, әріп тіркесінде 76-орында, бастапқы позицияда (ШЫ). Бұл әріп те сирекке жатады. Бұл әріп қосәріп (диграф) болғандықтан, арнайы пернетақтада орналастыру қажет болмады.

**Г (G)** әрпі – әріп жиілігі бойынша 22-орында, әріп тіркесінде 46-орында, бастапқы позицияда (ГЕ). Сондықтан үстіңгі шеткі жақта, алайда жиі кездесетін Е-ге жақын маңда орналастырылды.

**Ң (Ñ)** әрпі – әріп жиілігі бойынша 23-орында, яғни сирек, алайда әріп тіркесінде 28-орын тым сирек кездесетіндерге жатпайды. Сондықтан осы екі жағдай да ескеріліп, астыңғы шеткі, алайда жиі кездесетін Е-нің астына орналастырылды.

**Ж (J)** әрпі – әріп жиілігі бойынша 25-орында, әріп тіркесінде 35-орында бастапқы позицияда (ЖА). Тіркестегі жиілігі ескеріліп, жиі кездесетін Е-нің жанына екінші қатарға бір әріптен кейін шеткі орынға берілді.

**Ү (U)** әрпі – әріп жиілігі бойынша 26-орында, сирек кездесетіндерге жатады. Әріп тіркесінде де 87-орында бастапқы позицияда (ҮР) тұр. Тіркес жағынан да сирек кездесетін болғандықтан, оң жақ шеткі үстіңгі қатарға орналастырылды.

**Ө (Ó)** әрпі – әріп жиілігі бойынша 27-орында, яғни сирек, әріп тіркесінде 82-орында соңғы позицияда (КӨ). Сондықтан шеткі ортаңғы қатарға орналастырылды.

**Ү(Ü)** әрпі – әріп жиілігі бойынша 28-орында, әріп тіркесінде 117-орында. Сондықтан шеткі сол жақ үстіңгі қатарда.

**Ә (Á)** әрпі – әріп жиілігі бойынша 29-орында, әріп тіркесінде 128-орында, бастапқы позицияда (ӘН). Сондықтан сол жақ шеткі астыңғы қатарда.

**Х/х (H)** әріптері – өте сирек. Х – әріп жиілігінде 31-орында, h – 38-орында. Екеуінің жиілігі қосылатындықтан, ортаңғы қатарда, алайда сирек кездесетіндіктен шетке қарай орналастырылды.

**Ф (F)** әрпі – 33-орында, тіркесте 329-орында, бастапқы позицияда (ФИ), яғни өте сирек кездесетіндерге жатады. Сондықтан ең шетіндегі үстіңгі қатарға орналастырылды.

**В (V)** әрпі – 34-орында, әріп тіркесінде 359-орын, яғни өте сирек. Сондықтан F-ның астына орналастырылды.

**Ч (CH)** диграф құрамындағы С әрпі жоғарғы қатардағы сол жақ шеткі, орыстың Ё әрпі тұрған орынға берілді.

Осы әріптер орналастырылған пернетақта пилоттық жоба ретінде ұсынылып, пернетақтада теруге арналған арнайы компьютерлік бағдарлама жасалды. Жиілік талдау бойынша орналастырылған латын әріптері бірнеше пернетақта батырмасына жапсырылып, бұл пернетақта нұсқасының мәтін теруге ыңғайлы, тиімділігі уақыт пен сапасына қарай сынақтан өткізілуде.



тәуелді болып келеді. Сонымен, буын дегеніміз сөздің дыбысталу құрылымының негізгі бірліктерінің бірі деуге болады. Міне, осы бірлікте дауысты және дауыссыз фонемалардың аллофон түрлері жасалады деуге болады.

Қазақ тілінде, басқа да түркі тілдеріндегідей, буын бір немесе бірнеше фонемалардан тұрады және буын құрауға оның өзегі болып саналатын дауысты фонема міндетті түрде қатысуы керек.

А.Байтұрсынұлы сөйлеу мен сөйлем, сөйлем мен сөз, сөз бен буын, буын мен дыбыс желісін бір-бірінен туындатып, арасын үзбей өзара байланыстырып қарайды. Сөздерді буынға бөлудің жолын қарастыра келіп, «Буыншы әріп басқалардың ортасында қамауда тұрса, ол буынға «бітеу» деп ат қойындар; буын аяғы буыншы әріпке тірелсе, ол буынға «ашық» деп ат қойындар; буын буыншыдан басқа әріпке тірелсе, ол буынға «тұйық» деп ат қойындар» деп, буын түрлерін саралап береді [13, 344-345 бб.]. Сондағы «буыншы» әріптеріміз дауысты дыбыстар болып шығады. Буын құраудағы дауысты әріптердің мәнін «әлгі әріптер кіріспейтін буын» болмайды дейді.

Қ.Жұбановтың «Буын жігін қалай табуға болады?»<sup>1</sup> атты мақаласында буын түрлерінің өрнегін беруде дыбыс аттары мынадай қысқартылған пішінде жазылған: **Ды** – дауысты дыбыс; **Дз** – дауыссыз дыбыс; **сДЗ** – созымды дауыссыз дыбыс; **қДЗ** – қатаң дауыссыз.

Проф. Қ.Жұбанов буынның дыбыс құрамына қарай алты түрін көрсеткен [14, 361 б.].

1) **Ды** (Г) буыны немесе **жалаң** буын (жалғыз-ақ дауысты дыбыстан жасалады).

2) **Дз-Ды** (СГ) буыны немесе **ашық** буын.

3) **Ды-Дз** (ГС) буыны немесе **жеңіл тұйық** буын.

4) **Дз-Ды-Дз** (СГС) буыны немесе **жеңіл бітеу** буын.

5) **Ды-сДЗ-қДЗ** (ГСС) буыны немесе **ауыр тұйық** буын.

6) **Дз-Ды-сДЗ-қДЗ** (СГСС) буыны немесе **ауыр бітеу** буын.

Осы келтірілген буын түрлерінің алғашқы төртеуі Қ.Жұбановтың пайымдауынша: «... қазақ тілі буындарының негізгі түрлері» болып саналады.

Қ.Жұбанов жоғарыдағыдай буындардың моделі арқылы «буын жігін оңай, механик түрде табуға болады» – дейді. Келешекте қазақ тілінің буындық құрылымына статистикалық әдіс-тәсілдерді қолдану үшін арнайы компьютерлік алгоритмдер көмегімен буын түрін автоматты түрде тану және олардың жиілік сөздіктерін құрастыруда ғалымның «Буын жігін қалай табуға болады?» деген мақаласының берілген модельдерді қолданатын боламыз.

«Фонациялық ауаның қарқынымен кілт үзіліп шыққан бір немесе бірнеше дыбыс тобын буын дейміз. Демек, буын экспирация арқылы (өкпедегі ауаны сы-

<sup>1</sup> Мақала алғаш рет «Ауыл мұғалімі» журналының 1934 жылғы 2-санында жарияланған.

ртқа шығару арқылы) жасалады» – дей келіп, осы үрдісті проф. І.Кенесбаев та ұстанады, буын түрлерін ашық, тұйық және бітеу деп бөліп, олардың алты түрлі дыбыс құрамын көрсетеді [15, 260-261 бб.]. Буын сипатын, түрлерін, жігін, дыбыс құрамын кеңінен талдаған С.Мырзабеков «Буын — ... ауаның кілт үзілуінің не кедергіге ұшырауының нәтижесінде пайда болатын жеке (дауысты) дыбыс, не дыбыстар (дауысты, дауыссыз) тобы» деген анықтама береді [16, 103 б.].

Буын құбылысын сингармонизммен тікелей байланыстыра отырып, проф. Ә.Жүнісбек «Қазақ тіліндегі дыбыс түрленімі қазақ сөзінің сингармонизм заңдылығынан туындайды. Сингармонизм әуезі (жуан, жіңішке, еріндік, езулік) көрінетін тірек бірлік буын болып табылады», – деп теориялық қорытынды жасайды [17, 24-25 бб.].

Буынның, әсіресе, фонемалардың дистрибуциясына қатысты жақтары аса маңызды деуге болады. В.Б.Касевичтің сөзімен айтатын болсақ: «Слог есть способ синтагматической организации фонем; в слогe реализуются ограничения, налагаемые системой на фонемную синтактику. Данная функция слога – дистрибутивная – универсальна. Во всех языках существуют ограничения на сочетаемость фонем, наиболее типичны из них те, которые описывают сочетаемость фонем в рамках синтагматических групп, отождествимых с традиционными слогами» [18, 141-142.]. Ғалым көрсеткендей, қазақ тіліндегі сөздердің буын жігін ажыратуда фонемалардың тіркесімділік заңдылығын да қатаң ұстану қажет. Ал бұл қазақ тілінде сөздерді дұрыс буынға бөлуге септігін тигізеді. Сөздердегі буын жігінің дұрыс ажыратылуы, түштеп келгенде, қазақ сөздерін дұрыс жазуға мүмкіндік береді.

Қазақ тілін буын деңгейінде статистикалық әдіс-тәсілдермен зерттеуде мынадай мәселелердің шешімін табуға болады:

- 1) қазақ тіліндегі буын құрылымының статистикасын анықтау;
- 2) буындардың жалпы статистикасына талдау жүргізу;
- 3) буындардың құрылымдық типтерінің жалпы статистикасын анықтау;
- 4) сөзформалардың бастапқы буындарының статистикасына талдау жасау;
- 5) сөзформалардың ортаңғы шептегі буындарының статистикасын анықтау;
- 6) сөзформалардың соңғы буындарына статистикалық талдау;
- 7) сөзформалардың буындық «тереңдігіне» (ұзындығына) статистикалық талдау жасау;
- 8) буындар типтерінің дистрибуциялық құрылымына статистикалық талдау жасау.

Мұндай зерттеулердің теориялық тұрғыда маңыздылығы мынада:

- біріншіден, олар тілдің фонетикалық құрылысының ажыратылмас бөлігі болуында;
- екіншіден, фонотактика мен морфонология мәселелеріне қатыстылығында;

- үшіншіден, туыстас және туыстас емес тілдерді типологиялық тұрғыда зерттеуге материал болуында.
- төртіншіден, буын жігін дұрыс бөліп, соған сәйкес сөздерді дұрыс жазуға ықпал етуінде;
- қазақ жазуын латын қарпіне көшіруде емле ережелерін жазуда т.б.

## 2.2 Қазақ тіліндегі буындардың жиілік сөздігінің тарихы

Буынның жиілік сөздігін алу басқа тілдік бірліктерге қарағанда күрделірек. Өйткені қазақ тілінде сөздерді буынға бөлу айтылымда жақсы ажыратылғанмен, кирилл қарпінің әсерінен көптеген сөздерімізді буын жігін сақтамай жазу нормаға түскендіктен, ауызша буынды басқаша дыбыстаймыз, жазуда басқаша жазамыз. Сондықтан сөздерді буынға бөлу жазуға бағына бермейді. Бұл әсіресе жоғарыда сөз болған, орыс тілінен енген *ю, я, и*, сонымен қатар *у* дыбыстарына қатысты.

Қазақ тіл білімінде белгілі ғалым Қ.Бектаев қазақ буындарының жиілік сөздігін 1973 жылы шыққан «Қазақ тексінің статистикасы» атты жинақта береді. Буын жиілігін алудағы мәтін көркем әдебиет, ғылыми-техникалық, публицистикалық (газет) стильдер текстерінің әрқайсысынан 25 000 буынқолданыстан және екітомдық «Қазақ тілінің түсіндірме сөздігінен» 52 015 буын қолданыстары алынған. Бұл сөздік үш бағанда беріліп, бірінші бағанда буындар, екінші бағанда Сөздіктегі буын жиілігі, үшінші бағанда үш стильден алынған буындар жиілігі біріктіріліп беріледі. Берілген буындардың жалпы саны – 2891 болған. Ғалым бұл еңбекте буындарды сөз басы мен сөз соңына қатысты позициялық жиілігін де көрсеткен. Қазақ тіліндегі бір буынды сөздердің тізімін де буын жиілігін анықтауда жеке шығарып көрсеткен [19, 607.]. **13-кесте.**

**13-кесте. Қ.Бектаев көрсеткен ең жиі қолданылатын буындардың сөздігі**

Буын	Абсолютті жиілік	Буын	Абсолютті жиілік	Буын	Абсолютті жиілік	Буын	Абсолютті жиілік	Буын	Абсолютті жиілік
да	1625	лы	421	дык	249	тең	157	дей	118
де	1460	ген	420	лі	243	деп	153	кез	113
ды	1238	ци	419	тар	238	ка	153	пар	113
а	1190	лық	408	ру	237	кел	153	рыш	112
ла	1094	бір	405	ше	235	дам	152	нын	111



та	1072	нің	386	ші	225	өз	148	нін	110
о	1014	се	382	тер	221	жыл	144	дік	109
е	986	ал	367	ти	221	ри	143	ну	109
қа	978	ша	345	дан	214	шін	143	ліп	108
сы	977	ба	341	тін	210	біз	139	гін	107
ді	967	ме	340	ай	208	тік	138	қан	107
на	831	ре	332	ко	208	і	137	ли	107
ға	811	кі	331	қан	208	дер	135	бер	106
не	795	сі	327	ко	206	рек	135	мыз	105
те	768	бе	320	дің	202	сон	135	ғың	103
ты	768	бол	308	у	202	у	134	тіп	103
ның	742	дың	300	бас	199	ы	134	май	102
ке	690	сын	300	тан	196	жоқ	132	ңыз	102
ғы	659	со	296	бұл	195	зу	132	кет	101
я	646	дар	287	тың	192	би	131	нұқ	101
ма	623	мы	287	бар	190	нан	130	за	100
жа	617	кә	278	су	190	сін	130	міз	100
ра	609	лып	277	ор	181	сы	129	нен	100
мен	602	дай	274	кер	179	жи	126	рып	100
ған	577	тын	269	ту	179	му	126	ди	99
ті	562	ту	268	сан	178	он	124	ком	99
бо	561	функ	267	лу	177	тық	124	құ	99
ры	560	си	265	ні	176	дә	122	лын	99
са	543	ә	263	рін	171	жо	122	тап	99
рі	512	кен	262	ми	168	мә	121	лыс	98
ге	498	бұ	261	лай	165	па	119	мі	98
ле	477	қы	260	ар	164	йын	118	нық	97
гі	463	рын	259	же	163	ни	117	ол	97
шы	451	жә	256	тыр	162	лер	116	бай	96
лар	438	ө	256	зі	159	лік	116	ақ	94
ны	434	бі	449	ұ	159	қай	114	гер	94
қу	94	дім	70	вет	59	жағ	49	ныс	41
ло	94	жүр	70	ет	59	про	49	ро	41
фор	93	лел	70	нып	59	рай	49	сеп	41
қи	92	нал	70	пен	59	түс	49	кол	40
қу	92	сол	70	зе	58	лім	48	по	40
мес	91	теп	70	сіп	58	ой	48	тен	40

оң	91	жер	69	тас	58	хат	48	бұ	39
сіз	91	жі	69	тә	58	хов	48	тон	39
эр	90	йы	69	тұ	58	шық	48	шын	39
ріп	90	мек	69	ел	57	қыр	47	бә	38
ау	89	зы	68	сып	57	же	46	бу	38
мін	88	но	68	тұр	57	қын	46	нім	38
рақ	88	іс	68	лап	56	фе	46	рым	38
то	88	ва	67	рас	56	зір	45	сақ	38
рыс	87	ен	67	гра	55	лек	45	ық	38
ден	86	мұн	67	зін	55	ат	44	көз	37
сал	86	мыс	67	тау	55	гіз	44	қой	37
йін	84	ріс	67	ұл	55	жал	44	қор	37
тің	83	тал	67	күн	54	ин	44	ңы	37
бел	81	яқ	67	жақ	53	ман	44	бес	36
бе	80	көп	66	йік	53	сөз	44	кон	36
и	80	мей	66	кіш	53	деу	43	рен	36
жат	79	ріл	65	тып	53	ки	43	тұр	36
өт	79	ха	65	лис	52	мал	43	мон	35
бей	78	қар	64	об	52	нек	43	нам	35
жас	78	ңіз	64	ер	51	тай	43	пы	35
жұ	78	сет	64	жол	51	біл	42	тыс	35
мән	78	мо	63	жы	51	дау	42	уі	35
ын	77	тө	63	фи	51	жет	42	хоз	35
га	76	зақ	61	шек	51	мақ	42	айт	34
жү	76	пе	61	ши	51	нақ	42	аң	34
кін	75	үй	61	шо	51	пуб	42	дым	34
ер	75	ас	60	ғар	50	уш	42	йық	34
ең	74	ес	60	йым	50	гу	41	лен	34
рес	74	на	60	кер	50	ғыс	41	сыз	34
коз	73	пай	60	қол	50	йім	41	тор	34
қыт	73	бек	59	тек	50	мент	41	ең	33
рет	73	бет	59	тіл	50	мын	41	бақ	33

Қ.Бектаев ең жиі кездесетін буындар ретінде 64 буын тектегі барлық буынқолданыстың 50 пайызын, 200-і 75 пайызын, 370-і 87 пайызын қамтитынын айтады. Сөйтіп, автор, кітаптағы кестеде берілген буындарды қазақ тексіндегі ең жиі кездесетін буындар, олар барлық мәтіннің 87 пайызын, қалған буындар 13 пайызын құрайтынын айтады.

## 2.4 Қазақ тілі буындарының статистикалық сипаттамасы

Қазақ буындарының жиілік сөздігін қазіргі заманауи және қазақ тілінің барлық стильдерін қамтитын мәтіндерден алу мақсатында «Қазақ тіліндегі буындардың жиілік сөздігін» ұсынып отырмыз. Бұл жиілік сөздікті жасауда буын жігін автоматты жолмен ажырату қиын болғандықтан, Қолданбалы лингвистика бөлімінің серверде орналастырылған «Қазақ тілінің мәтіндер корпусынан» қазақ тілінің бес стилі бойынша жасалған 6 миллион сөзқолданыстан тұратын мәтіндер жинағынан әр стиль бойынша 10 000 сөзқолданыстан, барлығы – 50 000 сөзқолданыстан тұратын мәтін алынып, «қолдап» буынға бөлінді, яғни «Қазақ тілі буындарының жиілік сөздігін» құрастырудан бұрын көлемі 50 мың сөзқолданыстан тұратын мәтін бойынша сөзформалардың жиілік сөздігі құрастырылды. Бұл сөзформалардағы сөзтізбенің рет саны 32304 сөзформаға тең.

Терминдердегі буын жігі қазақ тілінің табиғатынан жат қолданыс болғандықтан, сөзтізбеден термин сөздер алынып тасталды.

Бұл «Қазақ тіліндегі буындардың жиілік сөздігі» дәстүрлі типі бойынша үш түрлі нұсқада жасалды: «Қазақ тіліндегі буындардың Әліпбилі жиілік сөздігі», «Қазақ тіліндегі буындардың жиілікті әліпбилі сөздігі», «Қазақ тіліндегі буындардың кері-әліпбилі жиілік сөздігі».

Төменде көрініс тапқан «Қазақ тілі буындарының жиілік сөздігінде» қазақ тіліндегі буындардың реттік саны 1225, олардың қайталана қолданылуы негізінде буындардың қосынды жиілігі 124063-ке тең.

Қазақ тілінде буынның үш түрі бар: 1) Ашық буын; 2) Тұйық буын; 3) Бітеу буын. Жиілік сөздікте буын түрлері шартты белгілермен таңбаланды, мысалы, қазақша: *ашық буын* – **АБ**, *тұйық буын* – **ТБ**, *бітеу буын* – **ББ** арқылы, ал *дауысты дыбыс* – **Ды**, *дауыссыз дыбыс* – **Дз** арқылы және олардың халықаралық шартты белгілері: *дауысты дыбыс* – **V**, *дауыссыз дыбыс* – **C** арқылы таңбаланды.

Енді қазақ тілі буындарының кері-әліпбилі жиілік сөздігінен шартты белгіленімдер бойынша буындардың түрлерін, яғни *ашық буын* – **АБ**, *тұйық буын* – **ТБ**, *бітеу буын* – **ББ** бойынша жеке-жеке қарастыруға болады. Ашық буынға қатысты кері-әліпбилі жиілік сөздігінен үзінді төменде **14-кестеде** көрініс тапты. **14-кесте.** Қазақ тіліндегі ашық буындардың кері-әліпбилі жиілік сөздігі

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1	<i>a/V/A</i>	1944	1,5669
2	<i>ә/V/A</i>	639	2,0820
3	<i>e/V/A</i>	1083	2,9550



Рет саны №	Буын	Абсолютті жілік	Бір топ буынның мәтінді қамту пайызы
38	<i>лә/C+V/A</i>	1	21,3722
39	<i>мә/C+V/A</i>	216	21,5463
40	<i>нә/C+V/A</i>	37	21,5761
41	<i>пә/C+V/A</i>	38	21,6068
42	<i>рә/C+V/A</i>	13	21,6172
43	<i>сә/C+V/A</i>	31	21,6422
44	<i>тә/C+V/A</i>	100	21,7228
45	<i>уә/C+V/A</i>	60	21,7712
46	<i>шә/C+V/A</i>	5	21,7752
47	<i>бе/C+V/A</i>	581	22,2435
48	<i>ге/C+V/A</i>	901	22,9698
49	<i>де/C+V/A</i>	2226	24,7640
50	<i>же/C+V/A</i>	364	25,0574
51	<i>зе/C+V/A</i>	97	25,1356
52	<i>ие/C+V/A</i>	35	25,1638
53	<i>йе/C+V/A</i>	243	25,3597
54	<i>ке/C+V/A</i>	1366	26,4607
55	<i>ле/C+V/A</i>	1058	27,3135
56	<i>ме/C+V/A</i>	705	27,8818
57	<i>не/C+V/A</i>	1366	28,9829
58	<i>ңе/C+V/A</i>	30	29,0070
59	<i>пе/C+V/A</i>	56	29,0522
60	<i>ре/C+V/A</i>	454	29,4181
61	<i>се/C+V/A</i>	638	29,9324
62	<i>те/C+V/A</i>	774	30,5562
63	<i>уе/C+V/A</i>	20	30,5724
64	<i>ше/C+V/A</i>	436	30,9238
65	<i>би/C+V/A</i>	142	31,0383
66	<i>ғи/C+V/A</i>	34	31,0657
67	<i>ди/C+V/A</i>	12	31,0753
68	<i>жи/C+V/A</i>	93	31,1503
69	<i>қи/C+V/A</i>	10	31,1584
70	<i>қи/C+V/A</i>	91	31,2317
71	<i>ли/C+V/A</i>	23	31,2503



Рет саны №	Буын	Абсолютті жілік	Бір топ буынның мәтінді қамту пайызы
106	<i>қу/C+V/A</i>	136	33,7893
107	<i>лу/C+V/A</i>	160	33,9183
108	<i>му/C+V/A</i>	7	33,9239
109	<i>ну/C+V/A</i>	71	33,9811
110	<i>ну/C+V/A</i>	1	33,9819
111	<i>ру/C+V/A</i>	478	34,3672
112	<i>су/C+V/A</i>	169	34,5034
113	<i>ту/C+V/A</i>	572	34,9645
114	<i>шу/C+V/A</i>	40	34,9967
115	<i>бу/C+V/A</i>	305	35,2426
116	<i>ғу/C+V/A</i>	6	35,2474
117	<i>ду/C+V/A</i>	33	35,2740
118	<i>жу/C+V/A</i>	263	35,4860
119	<i>зу/C+V/A</i>	4	35,4892
120	<i>йу/C+V/A</i>	6	35,4941
121	<i>қу/C+V/A</i>	534	35,9245
122	<i>лу/C+V/A</i>	6	35,9293
123	<i>му/C+V/A</i>	189	36,0817
124	<i>ну/C+V/A</i>	2	36,0833
125	<i>ну/C+V/A</i>	1	36,0841
126	<i>ру/C+V/A</i>	1	36,0849
127	<i>су/C+V/A</i>	92	36,1590
128	<i>ту/C+V/A</i>	237	36,3501
129	<i>шу/C+V/A</i>	18	36,3646
130	<i>бу/C+V/A</i>	126	36,4662
131	<i>ду/C+V/A</i>	84	36,5339
132	<i>жу/C+V/A</i>	177	36,6765
133	<i>ку/C+V/A</i>	225	36,8579
134	<i>му/C+V/A</i>	130	36,9627
135	<i>су/C+V/A</i>	22	36,9804
136	<i>ту/C+V/A</i>	277	37,2037
137	<i>шу/C+V/A</i>	3	37,2061
138	<i>бы/C+V/A</i>	148	37,3254
139	<i>ғы/C+V/A</i>	1285	38,3612





Жиілік сөздіктің реттік саны бойынша барлық **ашық буынға (АБ)** жата-тын бірліктер саны 173-ке тең, ал олар қайталана қолданыла келіп, мәтіннің 54,59 пайызын қамтитынына көз жеткізуге болады. **14-кесте.**

Кесте деректеріне сүйенетін болсақ, **ашық буындардың (АБ)** 11-і бір дауысты дыбыстан тұратын буындар, яғни **а, ә, е, и, о, ө, у, ұ, ү, ы, і** дауысты дыбыстар. Аталған буындар мәтін ішінде қайталана қолданылып, барлық буын түрлерінің 5,7 пайызын қамтиды. Аталған 11 дауысты дыбыстың ішінен ең жиі қолданғандары: **«а»** – 1944 рет, **«о»** – 1603 рет, **«е»** – 1083 рет.

Екі дыбысты ашық буындар мәтіннің қалған бөлігін – 48,89 пайызын құрайды.

Сол сияқты, **бітеу буын (ББ)** жайлы сөз ететін болсақ, олар кері-әліпбилі жиілік сөздікте ашық буынға қарағанда сандық жағынан айтарлықтай көп деуге болады. Мәселен, рет саны бойынша бітеу буынның реестрдегі саны 952-ге тең де, ал барлық буын бірліктерінің мәтін ішінде қайталана қолданылып, қазақ тілінің 5 стилі бойынша алынған мәтіннің 41 пайызын қамтиды. Ал бітеу буындардың екі моделі бойынша пайыздық көрсеткішін анықтайтын болсақ, **С+V+C+C** төрт әріпті бітеу буын – 0,24 %, **С+V+C** үш әріпті бітеу буын – 40,753%. Қазақ тіліндегі бітеу буындардың кері-әліпбилі жиілік сөздігінің үзіндісі **15-кестеде** берілді.

**15-кесте. Қазақ тіліндегі бітеу буындардың кері-әліпбилі жиілік сөздігінен үзінді**

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
174	<i>желк/С+V+C+C/Б</i>	2	54,5908
175	<i>бүлк/С+V+C+C/Б</i>	5	54,5948
176	<i>сұлк/С+V+C+C/Б</i>	2	54,5965
177	<i>сылқ/С+V+C+C/Б</i>	3	54,5989
178	<i>даңқ/С+V+C+C/Б</i>	4	54,6021
179	<i>жарк/С+V+C+C/Б</i>	2	54,6037
180	<i>жұрк/С+V+C+C/Б</i>	1	54,6045
181	<i>шырк/С+V+C+C/Б</i>	5	54,6085
182	<i>балп/С+V+C+C/Б</i>	2	54,6102
183	<i>тырп/С+V+C+C/Б</i>	1	54,6110
184	<i>дүре/С+V+C+C/Б</i>	2	54,6126
185	<i>мырс/С+V+C+C/Б</i>	6	54,6174
186	<i>ғайт/С+V+C+C/Б</i>	1	54,6182
187	<i>жайт/С+V+C+C/Б</i>	11	54,6271



Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
222	<i>ғырт/С+V+С+С/Б</i>	1	54,7931
223	<i>дырт/С+V+С+С/Б</i>	2	54,7947
224	<i>жсырт/С+V+С+С/Б</i>	3	54,7972
225	<i>йырт/С+V+С+С/Б</i>	1	54,7980
226	<i>қырт/С+V+С+С/Б</i>	3	54,8004
227	<i>пырт/С+V+С+С/Б</i>	2	54,8020
228	<i>сырт/С+V+С+С/Б</i>	16	54,8149
229	<i>тырт/С+V+С+С/Б</i>	1	54,8157
230	<i>уырт/С+V+С+С/Б</i>	3	54,8181
231	<i>шырт/С+V+С+С/Б</i>	1	54,8189
232	<i>гірт/С+V+С+С/Б</i>	1	54,8197
233	<i>дірт/С+V+С+С/Б</i>	1	54,8205
234	<i>кірт/С+V+С+С/Б</i>	3	54,8230
235	<i>сірт/С+V+С+С/Б</i>	1	54,8238
236	<i>тірт/С+V+С+С/Б</i>	1	54,8246
237	<i>қарш/С+V+С+С/Б</i>	4	54,8278
238	<i>кірш/С+V+С+С/Б</i>	2	54,8294
239	<i>жаб/С+V+С/Б</i>	4	54,8326
240	<i>раб/С+V+С/Б</i>	20	54,8487
241	<i>хаб/С+V+С/Б</i>	1	54,8496
242	<i>шаб/С+V+С/Б</i>	1	54,8504
243	<i>ғиб/С+V+С/Б</i>	1	54,8512
244	<i>қоб/С+V+С/Б</i>	1	54,8520
245	<i>сыб/С+V+С/Б</i>	4	54,8552
246	<i>мег/С+V+С/Б</i>	1	54,8560
247	<i>бөг/С+V+С/Б</i>	4	54,8592
248	<i>баг/С+V+С/Б</i>	87	54,9294
249	<i>даг/С+V+С/Б</i>	34	54,9568
250	<i>жаг/С+V+С/Б</i>	109	55,0446
---	-----	--	-----
1030	<i>быт/С+V+С/Б</i>	2	94,2279
1031	<i>ғыт/С+V+С/Б</i>	35	94,2561
1032	<i>зыт/С+V+С/Б</i>	1	94,2570
1033	<i>йыт/С+V+С/Б</i>	4	94,2602



Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1068	<i>беу/С+V+С/Б</i>	7	94,9735
1069	<i>зеу/С+V+С/Б</i>	19	94,9888
1070	<i>деу/С+V+С/Б</i>	41	95,0219
1071	<i>зеу/С+V+С/Б</i>	2	95,0235
1072	<i>йеу/С+V+С/Б</i>	2	95,0251
1073	<i>кеу/С+V+С/Б</i>	19	95,0404
1074	<i>леу/С+V+С/Б</i>	66	95,0936
1075	<i>меу/С+V+С/Б</i>	4	95,0968
1076	<i>неу/С+V+С/Б</i>	12	95,1065
1077	<i>ңеу/С+V+С/Б</i>	3	95,1089
1078	<i>пеу/С+V+С/Б</i>	3	95,1114
1079	<i>реу/С+V+С/Б</i>	27	95,1331
1080	<i>теу/С+V+С/Б</i>	58	95,1799
1081	<i>шеу/С+V+С/Б</i>	5	95,1839
1082	<i>йұу/С+V+С/Б</i>	22	95,2016
1083	<i>йүу/С+V+С/Б</i>	1	95,2024
1084	<i>йіу/С+V+С/Б</i>	3	95,2049
1085	<i>сах/С+V+С/Б</i>	1	95,2057
1086	<i>рux/С+V+С/Б</i>	8	95,2121
1087	<i>рүx/С+V+С/Б</i>	2	95,2137
1088	<i>сұx/С+V+С/Б</i>	10	95,2218
1089	<i>ғаш/С+V+С/Б</i>	68	95,2766
1090	<i>даш/С+V+С/Б</i>	1	95,2774
1091	<i>қаш/С+V+С/Б</i>	4	95,2806
1092	<i>лаш/С+V+С/Б</i>	28	95,3032
1093	<i>маш/С+V+С/Б</i>	3	95,3056
1094	<i>ңаш/С+V+С/Б</i>	4	95,3088
1095	<i>паш/С+V+С/Б</i>	1	95,3096
1096	<i>шаш/С+V+С/Б</i>	4	95,3129
1097	<i>кеш/С+V+С/Б</i>	20	95,3290
1098	<i>пеш/С+V+С/Б</i>	4	95,3322
1099	<i>шеш/С+V+С/Б</i>	1	95,3330
1100	<i>қош/С+V+С/Б</i>	3	95,3354
1101	<i>пош/С+V+С/Б</i>	8	95,3419



9 тұйық буынның ішінде ең жоғары жиілікте: *айт* – 97 рет, *ұлт* – 67 рет және *өйт* – 24 рет қолданыс тапқан екен.

Үш дыбысты тұйық буындарға қарағанда **екі дыбысты тұйық буындардың** (V/C) саны айтарлықтай жоғары, яғни ол 91 санына тең және қайталаана қолданылу нәтижесінде мәтінді қамту пайызы – 4,24%. Олардың ішінен ең жоғары қолданыстағылар (100-ден жоғарылары) деп мына 16 тұйық буынды атауға болады: *ал* – 587, *ай* – 360, *өз* – 264, *ар* – 240, *ол* – 221, *іс* – 211, *ел* – 210, *ау* – 183, *от* – 170, *ор* – 166, *ақ* – 153, *үй* – 149, *ең* – 143, *ат* – 138, *ас* – 121, *ал* – 106.

**16-кесте. Қазақ тіліндегі тұйық буындардың кері-әліпбилі жиілік сөздігі**

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1126	<i>айт/V+C+C/T</i>	97	95,6611
1127	<i>әйт/V+C+C/T</i>	11	95,6699
1128	<i>өйт/V+C+C/T</i>	24	95,6893
1129	<i>үйт/V+C+C/T</i>	3	95,6917
1130	<i>ұлт/V+C+C/T</i>	67	95,7457
1131	<i>ант/V+C+C/T</i>	1	95,7465
1132	<i>арт/V+C+C/T</i>	13	95,7570
1133	<i>ерт/V+C+C/T</i>	2	95,7586
1134	<i>өрт/V+C+C/T</i>	6	95,7634
1135	<i>аб/V+C/T</i>	3	95,7659
1136	<i>әб/V+C/T</i>	16	95,7788
1137	<i>еэ/V+C/T</i>	2	95,7804
1138	<i>аз/V+C/T</i>	2	95,7820
1139	<i>аз/V+C/T</i>	38	95,8126
1140	<i>ез/V+C/T</i>	1	95,8134
1141	<i>оз/V+C/T</i>	6	95,8183
1142	<i>өз/V+C/T</i>	264	96,0310
1143	<i>үз/V+C/T</i>	19	96,0464
1144	<i>ыз/V+C/T</i>	1	96,0472
1145	<i>із/V+C/T</i>	47	96,0851
1146	<i>ай/V+C/T</i>	360	96,3752
1147	<i>әй/V+C/T</i>	10	96,3833
1148	<i>ой/V+C/T</i>	95	96,4599
1149	<i>өй/V+C/T</i>	2	96,4615
1150	<i>үй/V+C/T</i>	17	96,4752





Рет саны №	Буын	Абсолютті жілік	Бір топ буынның мәтінді қамту пайызы
1185	<i>ұң/V+C/T</i>	2	98,2775
1186	<i>ың/V+C/T</i>	15	98,2896
1187	<i>аң/V+C/T</i>	29	98,3130
1188	<i>оң/V+C/T</i>	3	98,3154
1189	<i>өң/V+C/T</i>	2	98,3170
1190	<i>ып/V+C/T</i>	2	98,3186
1191	<i>ар/V+C/T</i>	240	98,5120
1192	<i>әр/V+C/T</i>	87	98,5822
1193	<i>ер/V+C/T</i>	83	98,6491
1194	<i>ор/V+C/T</i>	166	98,7829
1195	<i>өр/V+C/T</i>	25	98,8030
1196	<i>ұр/V+C/T</i>	28	98,8256
1197	<i>үр/V+C/T</i>	4	98,8288
1198	<i>ыр/V+C/T</i>	3	98,8312
1199	<i>ір/V+C/T</i>	7	98,8369
1200	<i>ас/V+C/T</i>	121	98,9344
1201	<i>әс/V+C/T</i>	93	99,0094
1202	<i>ес/V+C/T</i>	75	99,0698
1203	<i>ис/V+C/T</i>	3	99,0722
1204	<i>ос/V+C/T</i>	3	99,0747
1205	<i>өс/V+C/T</i>	13	99,0851
1206	<i>ұс/V+C/T</i>	54	99,1287
1207	<i>үс/V+C/T</i>	58	99,1754
1208	<i>ыс/V+C/T</i>	5	99,1794
1209	<i>іс/V+C/T</i>	211	99,3495
1210	<i>ам/V+C/T</i>	138	99,4608
1211	<i>әм/V+C/T</i>	1	99,4616
1212	<i>ем/V+C/T</i>	48	99,5003
1213	<i>ім/V+C/T</i>	11	99,5091
1214	<i>ом/V+C/T</i>	74	99,5688
1215	<i>өм/V+C/T</i>	170	99,7058
1216	<i>ау/V+C/T</i>	183	99,8533
1217	<i>әу/V+C/T</i>	10	99,8614
1218	<i>ауу/V+C/T</i>	21	99,8783



№ р/с	Буын/белгіленім	Абсолютті жілік	Бір топ буынның мәтінді қамту пайызы
12	<i>an/V+C/T</i>	29	2,6108
13	<i>ap/V+C/T</i>	240	2,8042
14	<i>apm/V+C+C/T</i>	13	2,8147
15	<i>ac/V+C/T</i>	121	2,9122
16	<i>am/V+C/T</i>	138	3,0235
17	<i>ay/V+C/T</i>	183	3,1710
18	<i>au/V+C/T</i>	21	3,1879
19	<i>a/A</i>	639	3,7030
20	<i>ab/V+C/T</i>	16	3,7159
21	<i>ay/V+C/T</i>	10	3,7239
22	<i>aym/V+C+C/T</i>	11	3,7328
23	<i>al/V+C/T</i>	106	3,8182
24	<i>am/V+C/T</i>	2	3,8198
25	<i>an/V+C/T</i>	50	3,8601
26	<i>an/V+C/T</i>	81	3,9254
27	<i>ap/V+C/T</i>	87	3,9956
28	<i>ac/V+C/T</i>	93	4,0705
29	<i>am/V+C/T</i>	1	4,0713
30	<i>ay/V+C/T</i>	10	4,0794
31	<i>au/V+C/T</i>	2	4,0810
32	<i>ba/C+V/A</i>	997	4,8846
33	<i>baз/C+V+C/Б</i>	87	4,9547
34	<i>баз/C+V+C/Б</i>	8	4,9612
35	<i>бай/C+V+C/Б</i>	218	5,1369
36	<i>бақ/C+V+C/Б</i>	67	5,1909
37	<i>бал/C+V+C/Б</i>	13	5,2014
38	<i>балл/C+V+C+C/Б</i>	2	5,2030
39	<i>бам/C+V+C/Б</i>	1	5,2038
40	<i>бан/C+V+C/Б</i>	26	5,2248
41	<i>бан/C+V+C/Б</i>	26	5,2457
42	<i>бар/C+V+C/Б</i>	324	5,5069

№ р/с	Буын/белгіленім	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
43	<i>бас/С+V+С/Б</i>	529	5,9333
44	<i>бат/С+V+С/Б</i>	25	5,9534
45	<i>бау/С+V+С/Б</i>	16	5,9663
46	<i>ба/С+V/А</i>	88	6,0373
47	<i>бәз/С+V+С/Б</i>	1	6,0381
48	<i>бәй/С+V+С/Б</i>	23	6,0566
49	<i>бал/С+V+С/Б</i>	4	6,0598
50	<i>бе/С+V/А</i>	581	6,5281
---	-----	---	----
1193	шыр/С+V+С/Б	10	98,7466
1194	шырк/С+V+С+С/Б	5	98,7506
1195	шырт/С+V+С+С/Б	1	98,7514
1196	шыт/С+V+С/Б	2	98,7531
1197	ші/С+V/А	580	99,2206
1198	шік/С+V+С/Б	27	99,2423
1199	шіл/С+V+С/Б	13	99,2528
1200	шім/С+V+С/Б	25	99,2730
1201	шін/С+V+С/Б	264	99,4857
1202	шін/С+V+С/Б	1	99,4866
1203	шіп/С+V+С/Б	13	99,4970
1204	шір/С+V+С/Б	36	99,5260
1205	шіт/С+V+С/Б	4	99,5293
1206	щы/С+V/А	8	99,5357
1207	ы/V/А	30	99,5599
1208	ыз/V+С/Т	1	99,5607
1209	ык/V+С/Т	29	99,5841
1210	ыл/V+С/Т	7	99,5897
1211	ым/V+С/Т	1	99,5905
1212	ын/V+С/Т	10	99,5986
1213	ың/V+С/Т	15	99,6107
1214	ып/V+С/Т	2	99,6123
1215	ыр/V+С/Т	3	99,6147
1216	ыс/V+С/Т	5	99,6187

№ р/с	Буын/белгіленім	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1217	ыш/V+C/T	3	99,6212
1218	і/V/A	138	99,7324
1219	із/V+C/T	47	99,7703
1220	іл/V+C/T	5	99,7743
1221	ім/V+C/T	2	99,7759
1222	ін/V+C/T	30	99,8001
1223	ір/V+C/T	7	99,8057
1224	іс/V+C/T	211	99,9758
1225	іш/V+C/T	30	100
Σ		124063	

Жоғары жиілікті зонаға абсолютті жиіліктері 1000-нан жоғары, реттік саны 1–26 аралығындағы буындардың қайталана қолданыла келіп, мәтіннің 30 пайызын қамтитын жиілік сөздік бөлігін жатқыздық және оны **бірінші топ** деп атаймыз. Аталған аралықта орналасқан 26 буынның 25-і ашық буын, ал тек 13-ші рет санында орналасқан «ган» бітеу буын ғана 1430 жиілікпен барлық мәтіннің 1,15 пайызын қамтып, жоғары жиілікті зонаға еніп тұр.

Енді буын жиіліктері 1000-нан төмен және 100-ден жоғары, яғни 100-реттен 1000 ретке дейін қолданылған ашық буындарды **екінші топқа** жатқызатын болсақ, олардың түрлерінің саны 73-ке тең және олар қайталана қолданылу нәтижесінде барлық мәтіннің 22,45 пайызын қамтиды екен.

**18-кесте. Қазақ тілі буындарының жиілікті әліпбилі сөздігінен үзінді**

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1	да/C+V/A	2522	2,0328
2	қа/C+V/A	2520	4,0641
3	де/C+V/A	2226	5,8583
4	а/V/A	1944	7,4253
5	та/C+V/A	1852	8,9180
6	ла/C+V/A	1793	10,3633
7	ды/C+V/A	1789	11,8053
8	о/V/A	1603	13,0974



Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
43	<i>қы/C+V/A</i>	644	42,2688
44	<i>ә/V/A</i>	639	42,7839
45	<i>лық/C+V+C/Б</i>	639	43,2990
46	<i>се/C+V/A</i>	638	43,8132
47	<i>тар/C+V+C/Б</i>	622	44,3146
48	<i>тын/C+V+C/Б</i>	613	44,8087
49	<i>кі/C+V/A</i>	604	45,2955
50	<i>ал/V+C/T</i>	587	45,7687
---	-----	---	-----
1194	<i>сұн/C+V+C/Б</i>	1	99,9750
1195	<i>сұт/C+V+C/Б</i>	1	99,9758
1196	<i>сірт/C+V+C+C/Б</i>	1	99,9766
1197	<i>сім/C+V+C/Б</i>	1	99,9774
1198	<i>тайт/C+V+C+C/Б</i>	1	99,9782
1199	<i>төм/C+V+C/Б</i>	1	99,9790
1200	<i>тұл/C+V+C/Б</i>	1	99,9798
1201	<i>тырн/C+V+C+C/Б</i>	1	99,9807
1202	<i>тырт/C+V+C+C/Б</i>	1	99,9815
1203	<i>тірт/C+V+C+C/Б</i>	1	99,9823
1204	<i>уант/C+V+C+C/Б</i>	1	99,9831
1205	<i>уас/C+V+C/Б</i>	1	99,9839
1206	<i>уес/C+V+C/Б</i>	1	99,9847
1207	<i>уың/C+V+C/Б</i>	1	99,9855
1208	<i>үйт/C+V+C/Б</i>	1	99,9863
1209	<i>хаб/C+V+C/Б</i>	1	99,9871
1210	<i>хақ/C+V+C/Б</i>	1	99,9879
1211	<i>хас/C+V+C/Б</i>	1	99,9887
1212	<i>шаб/C+V+C/Б</i>	1	99,9895
1213	<i>шалт/C+V+C+C/Б</i>	1	99,9903
1214	<i>шат/C+V+C/Б</i>	1	99,9911
1215	<i>шәл/C+V+C/Б</i>	1	99,9919
1216	<i>шел/C+V+C/Б</i>	1	99,9927
1217	<i>шеш/C+V+C/Б</i>	1	99,9936
1218	<i>шон/C+V+C/Б</i>	1	99,9944





### 3 ҚАЗАҚ ТІЛІНДЕГІ ҚОСЫМШАЛАРДЫҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ

#### 3.1 Қазақ тіліндегі түрленім қосымшалар мен түрленім формалар жиілік сөздігін құрастырудың қажеттілік негіздемесі

Қайсібір тілде болмасын сөздер жалғыз өздері тұрып, бір-бірімен байланыспаса, ой дұрыс жеткізілмейді. Сөз бен сөзді байланыстырудың құралы тілдегі грамматикалық бірліктер болып табылады. Грамматикалық бірліктер (қосымшалар, көмекші сөздер) түбір сөздерге жалғанып не тіркесіп келіп, сөздерді байланыстырады. Тілдегі лексемалар сияқты қосымшалардың да байланысу тәсілдері мен қолданыс аясы бар. Сөз түрлендіруші қосымшалар әдетте сөздерге талғамай жалғанады да, сөздерді түрлендіреді. Осындай сөз түрлендіруші қосымшалардың саны (варианттарын қосқанда) мен түрленім формалардың санын және олардың мәтіндердегі қолданыс жиілігін анықтаудың маңызы зор. Мұндай зерттеулер морфология, статистика салалары үшін ғана емес, орфография, емле мәселелерін шешуде де нақты тілдік дәйек болып табылады.

#### 3.2 Қазақ тіліндегі морфемалар (сөз таптары, қосымшалар) жиілік сөздігінің тарихы

Қазақ тілінде статистикалық талдаулар 70-жылдары басталғаннан бастап, сөз таптары, оның ішінде кейбір сөз таптары статистикалық тұрғыдан зерттеле бастады. Ә.Ахабаев пен Қ.Бектаев зат есімдердің, Ә.Ахабаев есімдіктердің морфологиялық құрылымына талдау жасаса, С.Мырзабеков туынды түбір туынды етістіктердің, Қ.Бектаев пен С.Мырзабеков етістіктің форма өзгерткіш аффикстеріне, А.Белботаев ғылыми-техникалық стильдегі сын есім аффикстеріне, А.Жұбанов пен Е.Жұбанов «Абай жолы» романындағы сын есімдердің лексика-морфологиялық формаларына статистикалық талдаулар жасады.

Ә.Ахабаев, Қ.Бектаев «Зат есім сөздердің статистика-морфологиялық сипаты» атты мақаласында газет мәтіндерінен сөз таптарының қайсысы жиі қолданылатынын анықтайды. Ғалымдардың көрсетуінше, ең жиі қолданылатын сөз табы – зат есім, екінші – етістік, үшінші орында – сын есімдер. Төртінші орында – шылау сөздер. Ал ең сирек кездесетіні – еліктеу сөздер мен одағайлар [20, 50.]. Бұл мақалада зат есімдердің жалғаулары мен күрделене қосылатын, яғни бірінен кейін бірі жалғанатын формалары туралы статистикалық мәліметтер берілген. Олар газет мәтіндері мен «Абай жолы»

романы мәтіндері бойынша жасалған. Сәйкесінше екі стильдегі жалғаулардың қолданысындағы айырмашылықтар айқындалады.

Ғалымдардың көрсетуінше, бір құрамды зат есім жалғауларынан тәуелдік, көптік, септік (ілік, барыс, жатыс) жалғаулары жиі, одан кейінгі жиілікте екі құрамды жалғаулар, сосын барып үш құрамды жалғаулар болып келеді [20, 68.].

Ә.Ахабаев «Есімдіктердің морфологиялық құрылымына статистикалық талдау» деп аталатын мақаласында есімдіктердің кездесу жиілігін жоғарыда аталған «Абай жолы» мен газет тілі бойынша салыстыра талдайды [21, 90 б.]. Сонымен қатар мақалада есімдіктерге жалғанған қосымшалардың да тізімдері мен жиіліктері көрсетіліп, оларға статистикалық талдаулар жасалған.

С.Мырзабеков «Туынды түбір етістіктердің құрылымына статистикалық талдау» атты мақаласында М.Әуезовтің «Абай жолы» романында қолданылған туынды түбір етістіктердің құрылымына сандық талдау жасалған. Ғалым «Абай жолы» романында негізгі түбір етістіктерден гөрі туынды түбір етістіктердің (80 пайызға жуығы) көп қолданылғанын анықтайды. Ал сөз тудыратын қосымшалардың ішінде жиі кездесетіні -ла – 1978, -да – 1243, -та – 1622. С.Мырзабеков бұл қосымшаның жиілігі туралы былай деп жазады: «Зат есімнен етістік тудыратын аффикстерді -ла/ле және олардың варианттарыңыз ойлау тіпті де мүмкін емес. Сонымен қатар, автордың көрсетуінше, қазақ тіліндегі туынды түбір етістіктердің дені етіс жұрнақтары (әсіресе өзгелік етіс) арқылы жасалған», – дей келіп, мұндай актив қолданыс етіс жұрнақтарын сөз тудырушы жұрнақтарға қойылатын критерийлерге сәйкес келмейтінін атап көрсетеді. С.Мырзабеков бұл мақалада сөз тудырушы жұрнақтардың жиілігіне тоқталған [22, 131 б.].

Қ.Бектаев, С.Мырзабеков «Етістіктің форма өзгерткіш аффикстерінің статистикалық талдауы» атты мақалада болымсыздық аффикстерінің қолданыс жиілігіне талдау жасап, етістіктердің негізінен болымды формада өте жиі қолданылатынын, ал болымсыз формада өте сирек қолданылатынын анықтайды. Сонымен қатар, көсемше, жақ, есімше, жедел өткен шақ, рай формаларының және олардың варианттарының жиілігін анықтап, статистикалық талдаулар жасайды [22, 117 б.]. Ғалымның көрсетуінше, бір зат есім сөзден (әже) 196 форма пайда болады. Есім сөздердің ішінен түрленуге бейім зат есімдерден кейін – есімдіктер дейді. «Абай жолы» романындағы ең жиі кездесетін етістік сөздер: бол – 209, отыр – 95, тұр – 89, жат – 69 түрлену формасымен кездесетінін айтады.

А.Белботаев «Ғылыми-техникалық стильдегі сын есім аффикстерінің қолданылуы туралы» деген мақалада осы стильде кездескен сөз таптарының, сонымен қатар сын есім тудырушы жұрнақтардың жиілігі туралы мәлімет беріледі. Басқа да ғалымдар көрсеткендей, кездесу жиілігі бойынша

реті: зат есім, етістік, сын есім, шылау, есімдік, үстеу, сан есім, одағай, еліктеу. Ғалым бұл мақалада сын есім тудырушы жұрнақтарға статистикалық талдау жасай отырып, сын есімнің кейбір жұрнақтарының статистикалық ақпараты бойынша олардың сөзжасам мен форма жасамға қатысына қарайтың пікірлер айтады [23, 230 б.].

А.Жұбанов «Статистика лексико-морфологических форм прилагательных в романе М.Ауезова «Абай жолы» («Путь Абая» деп аталатын мақалада «Абай жолы» романындағы сын есімдердің жасалу жолдарына қарай статистикалық талдау жасайды. Ғалымдардың көрсетуінше, сын есімдердің қосымша жалғану арқылы жасалуы өте жиі, одан кейінгі орында түбір тұлғада жиілігі жоғары, ал қос сөз бен біріккен сөз түрінде сирек, ал кірме аффикстер (-наз) ең төменгі жиілікте [24, 232 б.]. Сонымен қатар шырай тұлғаларының жиілігі де беріледі. Публицистика мәтінін де қоса қарастырғанда, авторлар сын есімнің -ғы/гі, -сыз/сіз, -лы/лі, -дай/дей қосымшаларының өте жиі кездесетінін, әсіресе, -ғы/гі жұрнағының жоғары жиілікте қолданылатынын анықтайды.

А.Жұбанов сөзформалар тудыруға арналған түрленім қосымшаларының тізімін сөздің соңғы дыбыстарына жалғануына қарай бөліп, топтастырып, зат есім мен етістіктің түрленім формаларының кестесін жасады [25, 195-221 бб.].

Көріп отырғанымыздай, қазақ тіл білімінде алғаш статистикалық зерттеулер жүргізіліп басталған кезеңде қазақ тіліндегі қосымшалар тізімі мен олардың мәтіндердегі қолданыс жиілігі туралы мәселелер басқа тілдік бірліктермен қатар зерттеу нысанына ілікті. Сөз таптары, сөз тудырушы жұрнақтар, сол сияқты түрленім қосымшалары тұрғысынан зерттеліп, біраз жайттардың басы ашылған болатын. Алайда көп жылдардан бері мұндай статистикалық зерттеулер сирек жүргізіліп, олардың нәтижелері лингвистикалық зерттеулерде басшылыққа алынбай кетті. Қазіргі жазу, әліпби мәселесінде ғылыми зерттеулер жүргізу үшін жоғарыдағыдай тілдік бірліктер туралы статистикалық мәліметтерді білу қажеттік тудырады. Сонымен қатар тілдік бірліктердің жиілігін шығару, ол статистика немесе әліпби, жазу мәселесінде ғана емес, сол тілдік бірліктердің жалпы сөздігін жасаудың да негізі болады.

### **3.3 Қазақ тіліндегі түрленім қосымшаларының статистикалық сипаттамасы**

Қазақ тіліндегі түрленім қосымшаларының, яғни форма тудырушы қосымшалардың жиілігі Қолданбалы лингвистика бөлімі құрастырған «Қазақ тілінің мәтіндер корпусы» деп аталатын 6 миллионнан тұратын бес стильден алынған корпус мәтіндерінен алынды [26.]. Олар корпусқа мәтіндер енгізуде-

гі теңгерімділік (репрезентативтілік) принципі бойынша қазақ әдеби тілінің әртүрлі стильдері мен жанрларынан таңдап алынды.

- **көркем проза** (М.Әуезов, Ә.Кекілбаев, Ж.Аймауытов, О.Бөкеев, Ғ.Мүсірепов, Б.Майлин шығармаларынан *1 миллион сөзқолданыс*);

- **поэзия** (М.Мақатаев, А.Байтұрсынұлы, Ж.Молдағалиев, К.Ахметов, М.Өтемісұлы, М.Жұмабаев, М.Шаханов, Т.Молдағалиев, Т.Айбергенов, Х.Ерғалиев, Ш.Құдайбердиев, Қ.Аманжолов және т.б. бойынша *1 миллион сөзқолданыс*);

- **ғылыми-гуманитарлық мәтіндер** (2012 – 2014 жылдар аралығында жарық көрген «Тілтаным» журналынан, филология саласы бойынша соңғы жылдары (2010 – 2015) жарық көрген ғылыми конференциялар жинақтарының мәтіндері, «Ұлы дала тұлғалары» сериясымен жарық көрген ғылыми жинақтар («Қ.Жұбанов», «І.Кеңесбаев», «А.Байтұрсынұлы»), т.б. *1 миллион сөзқолданыс*);

- **публицистикалық мәтіндер** («Ана тілі», «Егемен Қазақстан», «Жас Алаш», «Айқын», «Қазақ әдебиеті», «Жетісу», «Астана ақшамы», «Дала мен қала», «Ақиқат» т.б. газеттерден *1 миллион сөзқолданыс*);

- **ресми (іскери) стиль** мәтіндері (интернет сайттарынан іскери құжаттар *1 миллион сөзқолданыс*);

- **сөйлеу стилі** (интернет сайттарынан алынған сұхбаттар мәтіндерінен *1 миллион сөзқолданыс*) т.б.

Осы дереккөздер қазақ тіліндегі түрленім қосымшаларының жиілігін алуға негіз болды.

Мәтіндерден қосымшалар жиілігін алу үшін мәтін алдымен морфологиялық анализатордан өткізілді. Морфологиялық анализатор автоматты түрде мәтіндегі сөздерді түбір мен қосымшаға бөліп, содан соң лемматизацияланған (түбірі бөлінген) бөліктен бөлек екінші жағындағы сөзтүрленім формаларын арнайы қосымшалар кестесі бойынша бөлшектеп және кестеде берілген шартты грамматикалық белгіленімдері бойынша сипаттап шығады. Содан кейін бірдей қосымша бөліктерін грамматикалық сипаттамаларымен бірге топтастырып, бірдей формалардың санын шығарады. Ұсынылып отырған қосымшалар сөздігі осылай автоматты әдіспен алынып отыр. Автоматты әдіс болғандықтан, омонимдерге қатысты кейбір нақты емес жағдайлар кездесуі мүмкін. Алайда 95 пайыз дұрыс ақпарат береді деп ойлаймыз. Мұндай көлемді мәтіндерден жиілік алуды қолмен жүзеге асыру мүмкін емес.

«Қазақ тіліндегі қосымшалардың жиілік сөздігі» әдеттегідей үш типте жасалды: «Қазақ тіліндегі қосымшалардың әліпбилі жиілік сөздігі», Қазақ тіліндегі қосымшалардың жиілікті әліпбилі сөздігі», «Қазақ тіліндегі қосымшалардың кері-әліпбилі жиілік сөздігі».

Қазақ тіліндегі түрленім қосымшаларының статистикалық сипаттамасын анықтау үшін ең әуелі әрбір қосымшаға қолмен грамматикалық белгіленімдері қойылды. Біз қарастырған мәтіндерде түрленім қосымшалардың 26 түрі айқындалды.

**20-кесте. Қазақ тіліндегі түрленім қосымшалардың мәтінді қамту пайызы**

Рет саны	Қазақ тіліндегі түрленім қосымшалардың аталуы	Түрленім қосымшалардың шартты белгілері	Түрленім қосымшалардың нұсқа түрлерінің саны	Қосымшалардың абсолютті жиілігі	Қосымшалардың пайыздық салмағы	Бір топ қосымшалардың пайыздық салмағы
1.	Тәуелдік жалғау	ТЖ	14	1764469	26,20	26,20
2.	Табыс септік	ТС	5	706673	10,49	36,69
3.	Есімше	Е	26	630286	9,36	46,05
4.	Көсемше	К	14	624734	9,28	55,33
5.	Жіктік жалғау	ЖЖ	24	564927	8,62	63,95
6.	Барыс септік	БС	6	402036	5,97	69,92
7.	Көптік жалғау	КЖ	6	377366	5,60	75,52
8.	Жатыс септік	ЖС	6	368073	5,47	80,99
9.	Ілік септік	ІС	6	359374	5,34	86,33
10.	Тұйық етістік	ТЕ	1	220043	3,27	89,60
11.	Шығыс септік	ШС	6	168694	2,51	92,11
12.	Болымсыздық жұрнақ	БЖ	6	109664	1,63	93,74
13.	Қатыстық форма	ҚФ	6	85790	1,27	95,01
14.	Көмектес септік	КС	6	81411	1,21	96,22
15.	Салыстыру формасы	СФ	6	70395	1,05	97,27
16.	Шартты рай	ШР	2	60375	0,90	98,17
17.	Бұйрық рай	БР	24	58805	0,87	99,04
18.	Жоқтық форма	ЖФ	2	30871	0,23	99,27
19.	Субстантив форма	СУБ	4	11049	0,16	99,43
20.	Өтіну формасы	ӨФ	2	9406	0,14	99,57
21.	Сұраулық шылау	СШ	6	9377	0,14	99,71





реестрден ең аз бөлігін, яғни 10 реттік санын алып жатқан: *табыс септік, субстантив форма, тәуелдік форма, қалау рай, жоқтық форма, ұқсату формасы, шартты рай, өтіну формасы, тұйық етістік, атрибут-анықтауыш форма* түрлері жатады. Үшінші топқа жататын түрленім қосымшалар жиілік сөздігі реестрінің бар болғаны 12,69 пайызын қамтып жатқанына **21-кесте** бойынша көз жеткізуге болады.

**21-кесте. Қазақ тіліндегі түрленім қосымша вариант санының реестр бойын қамту пайызы**

Рет саны	Қазақ тіліндегі түрленім қосымшалардың аталуы	Түрленім қосымшалардың шартты белгілері	Түрленім қосымшаның вариант саны	Әр қосымша вариант санының реестрді қамту пайызы (%)	Бір топ қосымша вариант санының реестрді қамту (%)
1.	Есімше	Е	26	13,19797	13,19797
2.	Жіктік жалғау	ЖЖ	24	12,18274	25,38071
3.	Бұйрық рай	БР	24	12,18274	37,56345
4.	Тәуелдік жалғау	ТЖ	14	7,106599	44,67005
5.	Көсемше	К	14	7,106599	51,77665
6.	Шырай	Ш	10	5,076142	56,85279
7.	Қатыстық форма	ҚФ	6	3,045685	59,89848
8.	Барыс септік	БС	6	3,045685	62,94416
9.	Көптік жалғау	КЖ	6	3,045685	65,98985
10.	Жатыс септік	ЖС	6	3,045685	69,03553
11.	Ілік септік	ІС	6	3,045685	72,08122
12.	Шығыс септік	ШС	6	3,045685	75,12690
13.	Болымсыздық жұрнақ	БЖ	6	3,045685	78,17259
14.	Көмектес септік	КС	6	3,045685	81,21827
15.	Салыстыру формасы	СФ	6	3,045685	84,26396
16.	Сұраулық шылау	СШ	6	3,045685	87,30964
17.	Табыс септік	ТС	5	2,538071	89,84772
18.	Субстантив форма	СУБ	4	2,030457	91,87817
19.	Тәуелдік форма	ТФ	3	1,522843	93,40102
20.	Қалау рай	ҚР	3	1,522843	94,92386



21.	Жоқтық форма	ЖФ	2	1,015228	95,93909
22.	Ұқсағу формасы	ҰФ	2	1,015228	96,95431
23.	Шартғы рай	ШР	2	1,015228	97,96954
24.	Өтіну формасы	ӨФ	2	1,015228	98,98477
25.	Тұйық етістік	ТЕ	1	0,507614	99,49239
26.	Атрибут-анықтауыш форма	АТР	1	0,507614	100
	Σ		197	100	

Төменде **22, 23, 24-кестелерде** қосымшалардың әліпбилі жиілік сөздігі, жиілікті әліпбилі сөздігі және кері әліпбилі жиілік сөздігінен үзінділер берілді.

**22-кесте. Қазақ тіліндегі түрленім қосымшалардың әліпбилі жиілік сөздігі**

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
1	<i>а/К</i>	145956	2,1675
2	<i>ар/Е</i>	21061	2,4803
3	<i>атын/Е</i>	31693	2,9510
4	<i>ба/БЖ</i>	3705	3,0060
5	<i>бақ/Е</i>	904	3,0194
6	<i>бас/Е</i>	651	3,0291
7	<i>бе/БЖ</i>	11148	3,1946
8	<i>бек/Е</i>	1710	3,2200
9	<i>бен/КС</i>	2720	3,2604
10	<i>бенен/КС</i>	176	3,2630
11	<i>бес/Е</i>	623	3,2723
12	<i>бы/СШ</i>	40	3,2729
13	<i>быз/ЖЖ</i>	663	3,2827
14	<i>бын/ЖЖ</i>	34	3,2832
15	<i>бі/СШ</i>	2847	3,3255
16	<i>біз/ЖЖ</i>	363	3,3309
17	<i>бін/ЖЖ</i>	11	3,3311
18	<i>ге/БС</i>	98325	4,7912
19	<i>гей/ҚР</i>	211	4,7944



Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
54	<i>ер/Е</i>	13575	29,3637
55	<i>етін/Е</i>	22033	29,6909
56	<i>й/К</i>	101980	31,2053
57	<i>йтын/Е</i>	13035	31,3989
58	<i>йтін/Е</i>	7684	31,5130
59	<i>йық/БР</i>	1453	31,5346
60	<i>йын/БР</i>	4008	31,5941
61	<i>йік/БР</i>	993	31,6089
62	<i>йін/БР</i>	7315	31,7175
63	<i>к/ЖЖ</i>	38199	32,2848
64	<i>ке/БС</i>	26655	32,6806
65	<i>кей/ҚР</i>	8	32,6807
66	<i>кен/Е</i>	37854	33,2429
67	<i>кендіктен/К</i>	78	33,2440
68	<i>кі/ҚФ</i>	577	33,2526
69	<i>кін/БР</i>	1285	33,2717
70	<i>қ/ЖЖ</i>	12272	33,4539
71	<i>қа/БС</i>	44537	34,1153
72	<i>қан/Е</i>	43280	34,7581
73	<i>қаниша/К</i>	419	34,7643
74	<i>қы/ҚФ</i>	405	34,7703
75	<i>қын/БР</i>	1327	34,7900
76	<i>лар/КЖ</i>	100347	36,2802
77	<i>лау/Ш</i>	1451	36,3018
78	<i>лер/КЖ</i>	51815	37,0712
79	<i>леу/Ш</i>	1553	37,0943
80	<i>лық/СУБ</i>	4006	37,1538
81	<i>лік/СУБ</i>	2119	37,1853
82	<i>м/ТЖ</i>	85989	38,4622
83	<i>ма/БЖ</i>	51942	39,2336
84	<i>мақ/Е</i>	6455	39,3295
85	<i>мас/Е</i>	9110	39,4648
86	<i>ме/БЖ</i>	23052	39,8071
87	<i>мек/Е</i>	3077	39,8528



<b>Рет саны №</b>	<b>Қосымша</b>	<b>Абсолютті жиілік</b>	<b>Бір топ қосымшаның мәтінді қамту пайызы</b>
122	<i>п/К</i>	138231	58,4594
123	<i>па/БЖ</i>	10791	58,6197
124	<i>пақ/Е</i>	5418	58,7001
125	<i>пас/Е</i>	1254	58,7187
126	<i>пе/БЖ</i>	9026	58,8528
127	<i>пек/Е</i>	5026	58,9274
128	<i>пен/КС</i>	10676	59,0860
129	<i>пенен/КС</i>	761	59,0973
130	<i>пес/Е</i>	988	59,1119
131	<i>пы/СШ</i>	166	59,1144
132	<i>пыз/ЖЖ</i>	1127	59,1311
133	<i>пын/ЖЖ</i>	1989	59,1607
134	<i>пі/СШ</i>	5827	59,2472
135	<i>піз/ЖЖ</i>	417	59,2534
136	<i>пін/ЖЖ</i>	987	59,2681
137	<i>р/Е</i>	131190	61,2163
138	<i>рақ/Ш</i>	141	61,2184
139	<i>рек/Ш</i>	228	61,2218
140	<i>са/ШР</i>	36680	61,7665
141	<i>сай/БР</i>	228	61,7699
142	<i>се/ШР</i>	23695	62,1218
143	<i>сей/БР</i>	112	62,1234
144	<i>сы/ТЖ</i>	123146	63,9522
145	<i>сыз/ЖФ/ЖЖ</i>	21316	64,2688
146	<i>сыздар/ЖЖ</i>	571	64,2773
147	<i>сын/БР</i>	10274	64,4298
148	<i>сың/ЖЖ</i>	8172	64,5512
149	<i>сыңдар/ЖЖ</i>	626	64,5605
150	<i>сі/ТЖ</i>	50408	65,3091
151	<i>сіз/ЖФ/ЖЖ</i>	9555	65,4510
152	<i>сіздер/ЖЖ</i>	279	65,4551
153	<i>сін/БР</i>	6141	65,5463
154	<i>сің/ЖЖ</i>	6710	65,6460
155	<i>сіңдер/ЖЖ</i>	9122	65,7814



Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
190	<i>ім/ТЖ</i>	29618	98,5664
191	<i>іміз/ТЖ</i>	13849	98,7721
192	<i>ің/ТЖ</i>	35646	99,3014
193	<i>іңдер/БР</i>	632	99,3108
194	<i>іңіз/БР/ТЖ</i>	4660	99,3800
195	<i>іңіздер/БР</i>	139	99,3821
196	<i>ін/К</i>	41462	99,9978
197	<i>ірек/Ш</i>	147	100
Σ			6733759

23-кесте. Қазақ тіліндегі түрленім қосымшаларының жиілікті әліп-билі сөздігі

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
1	<i>і/ТЖ</i>	697483	10,3580
2	<i>ы/ТЖ</i>	557876	18,6428
3	<i>ды/ЖЖ/ТС</i>	310868	23,2593
4	<i>н/ТС</i>	297765	27,6813
5	<i>ді/ЖЖ/ТС</i>	222500	30,9855
6	<i>у/ТЕ</i>	220043	34,2533
7	<i>ған/Е</i>	156381	36,5756
8	<i>а/К</i>	145956	38,7432
9	<i>ты/ТС/ЖЖ</i>	142360	40,8573
10	<i>н/К</i>	138231	42,9101
11	<i>р/Е</i>	131190	44,8583
12	<i>сы/ТЖ</i>	123146	46,6871
13	<i>ң/ЖЖ/ТЖ</i>	116200	48,4128
14	<i>ға/БС</i>	108832	50,0290
15	<i>да/ЖС</i>	107278	51,6221
16	<i>ның/ІС</i>	103762	53,1630
17	<i>й/К</i>	101980	54,6775
18	<i>ген/Е</i>	100528	56,1704
19	<i>лар/КЖ</i>	100347	57,6606

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
20	<i>ge/BC</i>	98325	59,1208
21	<i>ып/К</i>	97778	60,5728
22	<i>е/К</i>	94592	61,9776
23	<i>ті/ТС/ЖЖ</i>	90765	63,3255
24	<i>м/ТЖ</i>	85989	64,6025
25	<i>де/ЖС</i>	80277	65,7946
26	<i>дың/ІС</i>	78055	66,9538
27	<i>на/BC</i>	74836	68,0651
28	<i>тар/КЖ</i>	72694	69,1447
29	<i>нда/ЖС</i>	70406	70,1903
30	<i>нің/ІС</i>	68220	71,2034
31	<i>нде/ЖС</i>	66735	72,1944
32	<i>мен/КС</i>	65171	73,1622
33	<i>дар/КЖ</i>	61996	74,0829
34	<i>дің/ІС</i>	59231	74,9625
35	<i>ма/БЖ</i>	51942	75,7339
36	<i>лер/КЖ</i>	51815	76,5034
37	<i>сі/ТЖ</i>	50408	77,2520
38	<i>не/BC</i>	48851	77,9774
39	<i>нан/ШС</i>	48340	78,6953
40	<i>тер/КЖ</i>	47584	79,4020
41	<i>қа/BC</i>	44537	80,0633
42	<i>қан/Е</i>	43280	80,7061
43	<i>дер/КЖ</i>	42930	81,3436
44	<i>ін/К</i>	41462	81,9593
45	<i>ғы/ҚФ</i>	39499	82,5459
46	<i>дан/ШС</i>	39257	83,1289
47	<i>к/ЖЖ</i>	38199	83,6962
48	<i>кен/Е</i>	37854	84,2583
49	<i>дай/СФ</i>	37186	84,8106
50	<i>са/ШР</i>	36680	85,3553
51	<i>гі/ҚФ</i>	36305	85,8944
52	<i>ың/ТЖ</i>	35988	86,4289
53	<i>ің/ТЖ</i>	35646	86,9583



Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
54	<i>ым/ТЖ</i>	34995	87,4779
55	<i>атын/Е</i>	31693	87,9486
56	<i>нен/ШС</i>	31070	88,4100
57	<i>та/ЖС</i>	30250	88,8592
58	<i>тың/ІС</i>	29885	89,3030
59	<i>ім/ТЖ</i>	29618	89,7429
60	<i>ке/БС</i>	26655	90,1387
61	<i>ден/ШС</i>	23945	90,4943
62	<i>се/ШР</i>	23695	90,8462
63	<i>ме/БЖ</i>	23052	91,1885
64	<i>етін/Е</i>	22033	91,5157
65	<i>сыз/ЖФ/ЖЖ</i>	21316	91,8323
66	<i>ар/Е</i>	21061	92,1451
67	<i>тің/ІС</i>	20221	92,4454
68	<i>мыз/ЖЖ/ТЖ</i>	18809	92,7247
69	<i>тан/ШС</i>	17546	92,9853
70	<i>ны/ТС</i>	17491	93,2450
71	<i>дей/СФ</i>	16669	93,4926
72	<i>мын/ЖЖ</i>	15563	93,7237
73	<i>ымыз/ТЖ</i>	14850	93,9442
74	<i>міз/ЖЖ/ТЖ</i>	14119	94,1539
75	<i>іміз/ТЖ</i>	13849	94,3595
76	<i>ер/Е</i>	13575	94,5611
77	<i>ниша/ҚФ/ҰФ</i>	13182	94,7569
78	<i>те/ЖС</i>	13127	94,9518
79	<i>йтын/Е</i>	13035	95,1454
80	<i>қ/ЖЖ</i>	12272	95,3277
81	<i>бе/БЖ</i>	11148	95,4932
82	<i>па/БЖ</i>	10791	95,6535
83	<i>пен/КС</i>	10676	95,8120
84	<i>сын/БР</i>	10274	95,9646
85	<i>мін/ЖЖ</i>	9983	96,1128
86	<i>сіз/ЖФ/ЖЖ</i>	9555	96,2547
87	<i>сіңдер/ЖЖ</i>	9122	96,3902



Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
122	<i>дік/СУБ</i>	2707	99,2404
123	<i>тау/Ш</i>	2288	99,2744
124	<i>дық/СУБ</i>	2217	99,3073
125	<i>лік/СУБ</i>	2119	99,3388
126	<i>пын/ЖЖ</i>	1989	99,3683
127	<i>менен/КС</i>	1907	99,3967
128	<i>бек/Е</i>	1710	99,4221
129	<i>улі/Е</i>	1559	99,4452
130	<i>леу/Ш</i>	1553	99,4683
131	<i>йық/БР</i>	1453	99,4899
132	<i>лау/Ш</i>	1451	99,5114
133	<i>қын/БР</i>	1327	99,5311
134	<i>кін/БР</i>	1285	99,5502
135	<i>ндей/СФ</i>	1279	99,5692
136	<i>ше/ҚФ/ҰФ</i>	1264	99,5880
137	<i>пас/Е</i>	1254	99,6066
138	<i>ғандықтан/К</i>	1215	99,6246
139	<i>пыз/ЖЖ</i>	1127	99,6414
140	<i>дау/Ш</i>	1065	99,6572
141	<i>йік/БР</i>	993	99,6719
142	<i>пес/Е</i>	988	99,6866
143	<i>пін/ЖЖ</i>	987	99,7013
144	<i>ыңдар/БР</i>	945	99,7153
145	<i>бақ/Е</i>	904	99,7287
146	<i>ғаниш/К</i>	804	99,7407
147	<i>ғалы/К</i>	787	99,7523
148	<i>пенен/КС</i>	761	99,7636
149	<i>дікі/ТФ</i>	743	99,7747
150	<i>ша/ҚФ/ҰФ</i>	684	99,7848
151	<i>быз/ЖЖ</i>	663	99,7947
152	<i>генше/К</i>	659	99,8045
153	<i>бас/Е</i>	651	99,8141
154	<i>тікі/ТФ</i>	632	99,8235
155	<i>іңдер/БР</i>	632	99,8329

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
156	<i>сыңдар/ЖЖ</i>	626	99,8422
157	<i>бес/Е</i>	623	99,8514
158	<i>нікі/ТФ</i>	611	99,8605
159	<i>кі/ҚФ</i>	577	99,8691
160	<i>сыздар/ЖЖ</i>	571	99,8776
161	<i>теу/Ш</i>	553	99,8858
162	<i>гелі/К</i>	512	99,8934
163	<i>ңдар/БР</i>	506	99,9009
164	<i>ғай/ҚР</i>	458	99,9077
165	<i>деу/Ш</i>	450	99,9144
166	<i>қанша/К</i>	419	99,9206
167	<i>піз/ЖЖ</i>	417	99,9268
168	<i>қы/ҚФ</i>	405	99,9328
169	<i>біз/ЖЖ</i>	363	99,9382
170	<i>мы/СШ</i>	354	99,9435
171	<i>ңдер/БР</i>	317	99,9482
172	<i>ырақ/Ш</i>	280	99,9523
173	<i>сіздер/ЖЖ</i>	279	99,9565
174	<i>гендіктен/К</i>	261	99,9603
175	<i>рек/Ш</i>	228	99,9637
176	<i>сай/БР</i>	228	99,9671
177	<i>гей/ҚР</i>	211	99,9703
178	<i>бенен/КС</i>	176	99,9729
179	<i>пы/СШ</i>	166	99,9753
180	<i>ңіздер/БР</i>	154	99,9776
181	<i>ірек/Ш</i>	147	99,9798
182	<i>мі/СШ</i>	143	99,9819
183	<i>рақ/Ш</i>	141	99,9840
184	<i>іңіздер/БР</i>	139	99,9861
185	<i>ңыздар/БР</i>	136	99,9881
186	<i>ғыр/БР</i>	130	99,9900
187	<i>ыңыздар/БР</i>	128	99,9919
188	<i>тік/АТР</i>	119	99,9937
189	<i>гір/БР</i>	113	99,9954

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
190	<i>сей/БР</i>	112	99,9970
191	<i>кендіктен/К</i>	78	99,9982
192	<i>бы/СШ</i>	40	99,9988
193	<i>бын/ЖЖ</i>	34	99,9993
194	<i>нсіз/ЖФ</i>	21	99,9996
195	<i>бін/ЖЖ</i>	11	99,9998
196	<i>кей/ҚР</i>	8	99,9999
197	<i>нсыз/ЖФ</i>	7	100
		6733759	

24-кесте. Қазақ тіліндегі түрленім қосымшаларының кері-әліпбилі жиілік сөздігі

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
1	<i>дік/СУБ</i>	2707	0,0402
2	<i>лік/СУБ</i>	2119	0,0717
3	<i>дық/СУБ</i>	2217	0,1046
4	<i>лық/СУБ</i>	4006	0,1641
5	<i>бек/Е</i>	1710	0,1895
6	<i>мек/Е</i>	3077	0,2352
7	<i>пек/Е</i>	5026	0,3098
8	<i>бақ/Е</i>	904	0,3232
9	<i>мақ/Е</i>	6455	0,4191
10	<i>пақ/Е</i>	5418	0,4996
11	<i>ған/Е</i>	156381	2,8219
12	<i>қан/Е</i>	43280	3,4646
13	<i>ген/Е</i>	100528	4,9575
14	<i>кен/Е</i>	37854	5,5197
15	<i>атын/Е</i>	31693	5,9903
16	<i>йтын/Е</i>	13035	6,1839
17	<i>етін/Е</i>	22033	6,5111
18	<i>йтін/Е</i>	7684	6,6252
19	<i>р/Е</i>	131190	8,5735
20	<i>ар/Е</i>	21061	8,8862
21	<i>ер/Е</i>	13575	9,0878

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
22	<i>бас/Е</i>	651	9,0975
23	<i>мас/Е</i>	9110	9,2328
24	<i>пас/Е</i>	1254	9,2514
25	<i>бес/Е</i>	623	9,2607
26	<i>мес/Е</i>	3729	9,3160
27	<i>пес/Е</i>	988	9,3307
28	<i>ушы/Е</i>	7480	9,4418
29	<i>улі/Е</i>	1559	9,4650
30	<i>уші/Е</i>	3988	9,5242
31	<i>у/ТЕ</i>	220043	12,7919
32	<i>ба/БЖ</i>	3705	12,8470
33	<i>ма/БЖ</i>	51942	13,6183
34	<i>па/БЖ</i>	10791	13,7786
35	<i>бе/БЖ</i>	11148	13,9441
36	<i>ме/БЖ</i>	23052	14,2865
37	<i>пе/БЖ</i>	9026	14,4205
38	<i>быз/ЖЖ</i>	663	14,4304
39	<i>пыз/ЖЖ</i>	1127	14,4471
40	<i>біз/ЖЖ</i>	363	14,4525
41	<i>піз/ЖЖ</i>	417	14,4587
42	<i>к/ЖЖ</i>	38199	15,0259
43	<i>қ/ЖЖ</i>	12272	15,2082
44	<i>бын/ЖЖ</i>	34	15,2087
45	<i>мын/ЖЖ</i>	15563	15,4398
46	<i>пын/ЖЖ</i>	1989	15,4694
47	<i>бін/ЖЖ</i>	11	15,4695
48	<i>мін/ЖЖ</i>	9983	15,6178
49	<i>пін/ЖЖ</i>	987	15,6324
50	<i>сың/ЖЖ</i>	8172	15,7538
51	<i>сің/ЖЖ</i>	6710	15,8534
52	<i>сыздар/ЖЖ</i>	571	15,8619
53	<i>сыңдар/ЖЖ</i>	626	15,8712
54	<i>сіздер/ЖЖ</i>	279	15,8754
55	<i>сіңдер/ЖЖ</i>	9122	16,0108

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
56	<i>ды/ТС/ЖЖ</i>	310868	20,6274
57	<i>ты/ТС/ЖЖ</i>	142360	22,7415
58	<i>ді/ТС/ЖЖ</i>	222500	26,0457
59	<i>ті/ТС/ЖЖ</i>	90765	27,3937
60	<i>сыз/ЖФ/ЖЖ</i>	21316	27,7102
61	<i>сіз/ЖФ/ЖЖ</i>	9555	27,8521
62	<i>дар/КЖ</i>	61996	28,7728
63	<i>лар/КЖ</i>	100347	30,2630
64	<i>тар/КЖ</i>	72694	31,3425
65	<i>дер/КЖ</i>	42930	31,9801
66	<i>лер/КЖ</i>	51815	32,7496
67	<i>тер/КЖ</i>	47584	33,4562
68	<i>мыз/ЖЖ/ТЖ</i>	18809	33,7355
69	<i>міз/ЖЖ/ТЖ</i>	14119	33,9452
70	<i>ң/ЖЖ/ТЖ</i>	116200	35,6708
71	<i>ымыз/ТЖ</i>	14850	35,8914
72	<i>іміз/ТЖ</i>	13849	36,0970
73	<i>м/ТЖ</i>	85989	37,3740
74	<i>ым/ТЖ</i>	34995	37,8937
75	<i>ім/ТЖ</i>	29618	38,3336
76	<i>ың/ТЖ</i>	35988	38,8680
77	<i>ің/ТЖ</i>	35646	39,3974
78	<i>ңыз/БР/ТЖ</i>	6445	39,4931
79	<i>ыңыз/БР/ТЖ</i>	4502	39,5599
80	<i>ңіз/БР/ТЖ</i>	4507	39,6269
81	<i>іңіз/БР/ТЖ</i>	4660	39,6961
82	<i>ы/ТЖ</i>	557876	47,9808
83	<i>сы/ТЖ</i>	123146	49,8096
84	<i>і/ТЖ</i>	697483	60,1676
85	<i>сі/ТЖ</i>	50408	60,9162
86	<i>а/К</i>	145956	63,0837
87	<i>ғанша/К</i>	804	63,0957
88	<i>қанша/К</i>	419	63,1019
89	<i>е/К</i>	94592	64,5066

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
90	<i>генше/К</i>	659	64,5164
91	<i>й/К</i>	101980	66,0309
92	<i>ғандықтан/К</i>	1215	66,0489
93	<i>гендіктен/К</i>	261	66,0528
94	<i>кендіктен/К</i>	78	66,0540
95	<i>п/К</i>	138231	68,1068
96	<i>ып/К</i>	97778	69,5588
97	<i>ін/К</i>	41462	70,1745
98	<i>ғалы/К</i>	787	70,1862
99	<i>гелі/К</i>	512	70,1938
100	<i>сай/БР</i>	228	70,1972
101	<i>сей/БР</i>	112	70,1989
102	<i>ййк/БР</i>	993	70,2136
103	<i>йық/БР</i>	1453	70,2352
104	<i>ғын/БР</i>	3892	70,2930
105	<i>йын/БР</i>	4008	70,3525
106	<i>қын/БР</i>	1327	70,3722
107	<i>сын/БР</i>	10274	70,5248
108	<i>гін/БР</i>	8520	70,6513
109	<i>йін/БР</i>	7315	70,7600
110	<i>кін/БР</i>	1285	70,7791
111	<i>сін/БР</i>	6141	70,8703
112	<i>ңыздар/БР</i>	136	70,8723
113	<i>ыңыздар/БР</i>	128	70,8742
114	<i>ңдар/БР</i>	506	70,8817
115	<i>ыңдар/БР</i>	945	70,8957
116	<i>ңіздер/БР</i>	154	70,8980
117	<i>іңіздер/БР</i>	139	70,9001
118	<i>ңдер/БР</i>	317	70,9048
119	<i>іңдер/БР</i>	632	70,9142
120	<i>ғыр/БР</i>	130	70,9161
121	<i>гір/БР</i>	113	70,9178
122	<i>ғай/ҚР</i>	458	70,9246
123	<i>гей/ҚР</i>	211	70,9277



Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
124	<i>кей/ҚР</i>	8	70,9278
125	<i>тік/АТР</i>	119	70,9296
126	<i>са/ШР</i>	36680	71,4743
127	<i>се/ШР</i>	23695	71,8262
128	<i>ға/БС</i>	108832	73,4424
129	<i>қа/БС</i>	44537	74,1038
130	<i>на/БС</i>	74836	75,2152
131	<i>ге/БС</i>	98325	76,6753
132	<i>ке/БС</i>	26655	77,0712
133	<i>не/БС</i>	48851	77,7967
134	<i>да/ЖС</i>	107278	79,3898
135	<i>нда/ЖС</i>	70406	80,4354
136	<i>та/ЖС</i>	30250	80,8846
137	<i>де/ЖС</i>	80277	82,0767
138	<i>нде/ЖС</i>	66735	83,0678
139	<i>те/ЖС</i>	13127	83,2627
140	<i>бен/КС</i>	2720	83,3031
141	<i>мен/КС</i>	65171	84,2710
142	<i>бенен/КС</i>	176	84,2736
143	<i>менен/КС</i>	1907	84,3019
144	<i>пенен/КС</i>	761	84,3132
145	<i>пен/КС</i>	10676	84,4717
146	<i>н/ТС</i>	297765	88,8937
147	<i>ны/ТС</i>	17491	89,1535
148	<i>ні/ТС</i>	8171	89,2748
149	<i>дан/ШС</i>	39257	89,8578
150	<i>нан/ШС</i>	48340	90,5757
151	<i>тан/ШС</i>	17546	90,8362
152	<i>ден/ШС</i>	23945	91,1918
153	<i>нен/ШС</i>	31070	91,6532
154	<i>тен/ШС</i>	8536	91,7800
155	<i>дың/ІС</i>	78055	92,9392
156	<i>ның/ІС</i>	103762	94,4801
157	<i>тың/ІС</i>	29885	94,9239

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
158	<i>дің/IC</i>	59231	95,8035
159	<i>нің/IC</i>	68220	96,8166
160	<i>тің/IC</i>	20221	97,1169
161	<i>нсыз/ЖФ</i>	7	97,1170
162	<i>нсіз/ЖФ</i>	21	97,1173
163	<i>ғы/ҚФ</i>	39499	97,7039
164	<i>қы/ҚФ</i>	405	97,7099
165	<i>гі/ҚФ</i>	36305	98,2491
166	<i>кі/ҚФ</i>	577	98,2576
167	<i>шы/ӨФ</i>	6185	98,3495
168	<i>ші/ӨФ</i>	3221	98,3973
169	<i>дай/СФ</i>	37186	98,9495
170	<i>ндай/СФ</i>	3027	98,9945
171	<i>тай/СФ</i>	8578	99,1219
172	<i>дей/СФ</i>	16669	99,3694
173	<i>ндай/СФ</i>	1279	99,3884
174	<i>тей/СФ</i>	3656	99,4427
175	<i>дікі/ТФ</i>	743	99,4538
176	<i>нікі/ТФ</i>	611	99,4628
177	<i>тікі/ТФ</i>	632	99,4722
178	<i>ша/ҚФ/ҰФ</i>	684	99,4824
179	<i>нша/ҚФ/ҰФ</i>	13182	99,6781
180	<i>ше/ҚФ/ҰФ</i>	1264	99,6969
181	<i>нше/ҚФ/ҰФ</i>	2877	99,7396
182	<i>рек/Ш</i>	228	99,7430
183	<i>ірек/Ш</i>	147	99,7452
184	<i>рақ/Ш</i>	141	99,7473
185	<i>ырақ/Ш</i>	280	99,7514
186	<i>дау/Ш</i>	1065	99,7673
187	<i>лау/Ш</i>	1451	99,7888
188	<i>тау/Ш</i>	2288	99,8228
189	<i>деу/Ш</i>	450	99,8295
190	<i>леу/Ш</i>	1553	99,8525
191	<i>теу/Ш</i>	553	99,8607

Рет саны №	Қосымша	Абсолютті жиілік	Бір топ қосымшаның мәтінді қамту пайызы
192	<i>бы/СШ</i>	40	99,8613
193	<i>мы/СШ</i>	354	99,8666
194	<i>пы/СШ</i>	166	99,8691
195	<i>бі/СШ</i>	2847	99,9113
196	<i>мі/СШ</i>	143	99,9135
197	<i>ні/СШ</i>	5827	100
Σ		6733759	

Сонымен, жоғарыдағы статистикалық мәліметтерді қорытындыласақ: қосымша атауларының 12-сі жоғары жиілікте кездеседі. Олар: *тәуелдік жалғау, табыс септік, есімше, көсемше, жіктік жалғау, барыс септік, көптік жалғау, жатыс септік, ілік септік, тұйық етістік, шығыс септік, болымсыздық жұрнақ* түрлері қайталана қолданыла келіп, мәтіннің 93,74 пайызын қамтыған. Осы қосымша атаулары ішінде жиілігі жағынан *тәуелдік жалғау, табыс септік, есімше, көсемше* түрлері мәтіннің 55,33 пайызын қамтып, қазақ тілінде ең жиі кездесін қосымшалар екендігін көрсетеді.

Берілген түрленім қосымшалардың статистикалық мәліметтерінен мынадай лингвистикалық тұжырым жасауға болады. Түрленім қосымшалар ішінен қолданыста тәуелдік жалғауы ең жиі кездеседі екен. 6 миллион сөзқолданыстан тұратын қазақ тілінің 5 стилінен алынған мәтіндерде тәуелдік жалғаудың көп кездесуі, біріншіден, 1-2 жақ тәуелдік жалғауының жіктік жалғауының 1-2 жағымен омоним болып қалуымен байланысты, яғни тәуелдік жалғау жиілігі жіктік жалғаумен де қосылып, араласып кетуі мүмкін. Екіншіден, қазақ тілінде меншіктік ұғымның қарым-қатынас кезінде жиі қолданылатынын көрсетеді.

Сол сияқты табыс септігінің де жиілігі кейде 3 жақ жіктік жалғауымен омоним болып, араласып кетуі мүмкін. Мысалы, *кітапты, айтты*. Сөзформаларды автоматты түрде түбір мен қосымшаға бөліп алғаннан кейін екі бірдей қосымшаны анализатор омоним қосымша ретінде көрсетеді. Екіншіден, табыс септігінің жоғары жиілікті көрсетуі қолданыста, сөйлеуде объектілік қатынастың көп орын алатындығының айғағы.

Ал есімше, көсемше формаларының жиі кездесуі олардың формаларының көп болуымен байланысты әрі сөзді тиянақтаушы, дәнекерлеуші қызметімен байланысты, яғни предикаттық қатынасты орнатушы сөйлемде етістік десек, сол етістіктерді грамматикалық қатынасқа түсіретін тұлғалар есімше, көсемше формалары болып табылады.

Өте жиі кездесетіндерден басқа жиі және орта жиіліктегі түрленім қосымшалар туралы жеке-жеке тоқталмай, ендігі кезекте тек өте сирек кездесетін қосымшалардың не себепті сирек кездесуі туралы мынадай тұжырым

жасауға болар еді. Бұл қосымшалар: **субстантив форма, өтіну формасы, сұраулық шылау, ұқсату формасы, шырай, тәуелдік форма, қалау рай, атрибут-анықтауыш форма.**

Сирек кездесетіндігі анықталған бұл түрленім қосымшалар негізінен жаңа ағаулар берілген субстантив, өтіну, ұқсату, атрибут-анықтауыш формалар. Субстантивтендіру бұл көбінесе сөйлеу бірлігі болып табылады, әрі көбіне нормаға жата бермейді. Жиілік сөздік жасауға алынған мәтіндер көпшілігі жазба мәтіндер болғандықтан, субстантив, атрибут-анықтауыш формалардың сирек кездесуі заңды. Сондай-ақ *-шы/ші* қосымшасының төменгі жиіліктегі зонада тұруы оның бұйрық райға көп жалғана бермейтіндігін көрсетеді, яғни қолданыста бұйрықтық сөйлемдердің сыпайы түрде берілуі сирек болып тұр.

Сұраулық шылау деп жеке тұрғандағы емес, сөз жігіндегі *мы/мі/бы/бі, ны/ні* формалары алынып отыр. Сұраулық шылаудың осылай (айтамысың, барамысың) кірігіп қолданылуы аз санды көрсетеді, демек, сұраулық шылаулар көбінесе сөйлем соңында сөзден бөлек жиі қолданылады. Тәуелдік форманың қолданысы төменгі жиілікті көрсетіп отыр, яғни меншіктік қатынас қазақ тілінде көбінесе тәуелдік жалғауларының еншісінде. Рай түрлерінің ішінде қалау рай төменгі жиілікті зонада. Бұл қазақ тілінде модальділікті білдіруде қалау мәнінің бұйрық, шарттылық қатынасқа құрылған сөйлемдерден гөрі сирек қолданылатынын көрсетеді. Алайда бұл сандық мәлімет екінші жағынан сөз құрамын автоматты түрде ажыратуда жатыс септігінен кейін келетін *-ғы/гі/қы/кі* тұлғаларымен шатасуы да мүмкін.

Шырай тұлғаларының сирек кездесуі оның сөзжасамға жақын қызметімен де байланысты, яғни шырай тұлғалары сапалық сын есімдерге ғана жалғанатын, жаппай грамматикалық абстракция жасамайтын категория.

**21-кесте** бойынша тұжырым жасасақ, қазақ тіліндегі түрленім қосымшалардан **есімше, жіктік жалғау, бұйрық рай, тәуелдік жалғау, көсемше, шырай** формалары варианттары бойынша реестр бойын қамту пайызы жағынан жоғары жиілікті көрсетті. Мәселен, есімшенің – 26, жіктің жалғауының – 24, бұйрық райдың – 24, тәуелдік жалғаудың – 14, көсемшенің – 14, шырайдың – 10 вариант-нұсқасы кездесті. Бұл түрленім қосымшалар жиілік сөздік реестрінің 56,85 пайызын, яғни реестрлік рет санының жартысынан көбін қамтиды екен. Ал басқа түрленім қосымшалардың варианттар саны 6-дан төмен.

### 3. 4 Қазақ тіліндегі түрленім формалардың статистикалық сипаттамасы

«Қазақ тіліндегі қосымшалардың жиілікті әліпбилі сөздігінде» жалаң қосымшалардың жиілігі берілген болса, келесі статистикалық мәліметтер қазақ тіліндегі түрленім формалары бойынша жасалды. Яғни **түрленім формалар дегеніміз** – сөздің түбіріне (негізгі және туынды) жалғанатын қосымшалардың жеке немесе бірінің үстіне бірі жалғануындағы сөздің түрленуі.

Бұл сөздік те жоғарыда аталған 6 миллиондық корпус мәтінінен автоматты жолмен алынды. «Қазақ тіліндегі түрленім формаларының әліпбилі жиілік сөздігі», Қазақ тіліндегі түрленім формаларының жиілікті әліпбилі сөздігі», «Қазақ тіліндегі түрленім формаларының кері-әліпбилі жиілік сөздігі» атты үш типтен тұрады.

Қазақ тіліндегі түрленім формалардың статистикалық сипаттамасы қазақ тілінің түрленім формаларының кері-әліпбилі және жиілікті әліпбилі жиілік сөздіктерінің статистикалық мәліметтеріне қатысты айтылады. Ол үшін ең алдымен сөздік құрастырылатын мәтіндерге немесе сөзформалардың жиілік сөздіктеріндегі бірліктерге түрленім формаларды ажыратуға арналған шартты белгіленімдерді (разметкаларды) автоматты түрде немесе қол күшімен енгізілуі қажет. Төменде, біз ұстаным еткен шартты белгіленімдер түрлерімен **25-кесте** бойынша танысуға болады.

**25-кесте. Түрленім формаларды ажыратуға арналған шартты белгіленімдер**

Рет саны	Түрленім формалардың атаулары	Түрленім формалардың белгіленімдері	Рет саны	Түрленім формалардың атаулары	Түрленім формалардың белгіленімдері
1.	Субстантив форма	<b>СУБ</b>	15.	Барыс септік	<b>БС</b>
2.	Қарғыс форма	<b>ҚАРҒ</b>	16.	Жатыс септік	<b>ЖС</b>
3.	Есімше	<b>Е</b>	17.	Көмектес септік	<b>КС</b>
4.	Тұйық етістік	<b>ТЕ</b>	18.	Табыс септік	<b>ТС</b>
5.	Болымсыздық жұрнақ	<b>БЖ</b>	19.	Шығыс септік	<b>ШС</b>
6.	Жіктік жалғау	<b>ЖЖ</b>	20.	Ілік септік	<b>ІС</b>
7.	Көптік жалғау	<b>КЖ</b>	21.	Жоқтық форма	<b>ЖФ</b>
8.	Тәуелдік жалғау	<b>ТЖ</b>	22.	Қатыстық форма	<b>ҚФ</b>
9.	Функционалды жұрнақ	<b>ФЖ</b>	23.	Өтіну форма	<b>ӨФ</b>
10.	Көсемше	<b>К</b>	24.	Салыстыру форма	<b>СФ</b>
11.	Бұйрық рай	<b>БР</b>	25.	Тәуелдік форма	<b>ТФ</b>
12.	Қалау рай	<b>ҚР</b>	26.	Шырай	<b>Ш</b>
13.	Атрибут анықтауыш форма	<b>АТР</b>	27.	Жедел өткен шақ	<b>ЖӨШ</b>
14.	Шартты рай	<b>ШР</b>			

25-кестеде берілген тізім алдыңғы тақырыптардағы түрленім қосымшалардың тізімінен аздап ерекшеленеді. Түрленім формаларда *-ша/ше* қо-

сымшасын барлық жағдайда ұқсату формасы деп тануға келмейтін болғандықтан, біз көпқызметтігін ескеріп, функционалды жұрнақ деп атап отырмыз. Сондай-ақ бұл тізімге *-ғыр/гір/қыр/кір* қосымшаларын қарғыс форма деп енгізіп отырмыз. Жеке қосымшалардың жиілік сөздігінде бұл форма кездеспеген болатын, сондықтан 20-кестедегі тізімге енбеді.

Қазақ тілінің түрленім формаларының кері-әліпбилі және жиілікті әліпбилі жиілік сөздіктерінің (**26-кесте** және **27-кесте**) статистикалық мәліметтерін пайдалана отырып, біз **25-кестедегі** саны 27-ге тең шартты белгіленімдер негізіндегі түрленім формалардың реестр бойы мен мәтін көлемін қамту пайыздарына қатысты келесі статистикалық сипаттаманы оқырман қауымына ұсынып отырмыз. Ескерте кететін жайт, ол сөз соңына қойылатын түрленім форма белгіленімдері («/») жақшадан кейін орын алған) бірнешеу болған жағдайда статистикалық сипаттамаға тізбектің соңғы белгіленімі ғана есепке алынды. Мысалы, кері-әліпбилі жиілік сөздігі реестрінде орын алған *ма/йтын/дай/мын/БЖ/Е/СФ/ЖЖ* бірліктің ең соңғы реттегі жіктік жалғау белгіленімі, яғни «ЖЖ» ғана статистикалық сипаттамада сөз болады (**26-кесте** және **27-кестелерді** қара).

**26-кесте. Қазақ тілінің түрленім формалары кері-әліпбилі жиілік сөздігінен үзінді**

Рет саны №	Түрленім форма	Абсолютті жиілік	Бір топ түрленім формалардың мәтінді қамту пайызы
1	<i>бе/ген/дік/БЖ/Е/СУБ</i>	1	0,00002
2	<i>ме/ген/дік/БЖ/Е/СУБ</i>	6	0,00015
3	<i>не/ген/дік/БЖ/Е/СУБ</i>	3	0,00021
4	<i>ба/ған/дық/БЖ/Е/СУБ</i>	8	0,00038
5	<i>ма/ған/дық/БЖ/Е/СУБ</i>	10	0,00059
-----	-----	-----	-----
12	<i>ба/ғыр/БЖ/ҚАРҒ</i>	3	0,00539
13	<i>ма/ғыр/БЖ/ҚАРҒ</i>	17	0,00575
14	<i>па/ғыр/БЖ/ҚАРҒ</i>	1	0,00577
-----	-----	-----	-----
17	<i>ғыр/ҚАРҒ</i>	109	0,00818
18	<i>гір/ҚАРҒ</i>	108	0,01045
-----	-----	-----	-----

Рет саны №	Түрленім форма	Абсолютті жілік	Бір топ түрленім формалардың мәтінді қамту пайызы
20	<i>ма/ған/БЖ/Е</i>	4746	0,11048
21	<i>ме/ген/БЖ/Е</i>	2420	0,16147
22	<i>пе/ген/БЖ/Е</i>	853	0,17945
23	<i>ба/йтын/БЖ/Е</i>	227	0,18423
24	<i>ма/йтын/БЖ/Е</i>	2617	0,23938
-----	-----	-----	-----
36	<i>ған/Е</i>	115647	2,86753
37	<i>қан/Е</i>	29037	3,47939
38	<i>ген/Е</i>	70834	4,97199
39	<i>кен/Е</i>	21122	5,41706
40	<i>атын/Е</i>	26054	5,96607
-----	-----	-----	-----
60	<i>ба/у/БЖ/ТЕ</i>	131	8,30151
61	<i>ма/у/БЖ/ТЕ</i>	695	8,31616
62	<i>па/у/БЖ/ТЕ</i>	213	8,32065
63	<i>бе/у/БЖ/ТЕ</i>	72	8,32217
64	<i>ме/у/БЖ/ТЕ</i>	299	8,32847
-----	-----	-----	-----
89	<i>ме/ген/сіздер/БЖ/Е/ЖЖ</i>	16	10,95169
90	<i>ме/ген/сіңдер/БЖ/Е/ЖЖ</i>	4	10,95177
91	<i>ған/быз/Е/ЖЖ</i>	273	10,95753
-----	-----	-----	-----
211	<i>лар/мыз/КЖ/ЖЖ</i>	57	11,12231
212	<i>тар/мыз/КЖ/ЖЖ</i>	56	11,12349
213	<i>дар/сыз/КЖ/ЖЖ</i>	16	11,12382
214	<i>лар/сыз/КЖ/ЖЖ</i>	44	11,12475
215	<i>тар/сыз/КЖ/ЖЖ</i>	4294	11,21523

Рет саны №	Түрленім форма	Абсолютті жилік	Бір топ түрленім формалардың мәтінді қамту пайызы
----	-----	-----	-----
471	<i>ма/ған/дай/мыз/БЖ/Е/СФ/ ЖЖ</i>	1	16,23891
472	<i>ма/йтын/дай/мын/БЖ/Е/ СФ/ЖЖ</i>	1	16,23893
473	<i>ма/йтын/дай/сын/БЖ/Е/ СФ/ЖЖ</i>	2	16,23897
474	<i>ған/дай/мыз/Е/СФ/ЖЖ</i>	14	16,23927
----	-----	-----	-----
638	<i>ы/нда/ғы/лар/ТЖ/ЖС/ҚФ/ КЖ</i>	5	22,73141
639	<i>де/гі/лер/ЖС/ҚФ/КЖ</i>	29	22,73203
640	<i>те/гі/лер/ЖС/ҚФ/КЖ</i>	73	22,73356
641	<i>ма/ған/дыз/ымыз/БЖ/Е/ СУБ/ТЖ</i>	1	22,73358
642	<i>ма/йтын/дыз/ым/БЖ/Е/ СУБ/ТЖ</i>	2	22,73363
----	-----	-----	-----
987	<i>ген/і/нше/Е/ТЖ/ФЖ</i>	51	41,81067
988	<i>у/ымыз/ша/ТЕ/ТЖ/ФЖ</i>	39	41,81149
989	<i>у/ым/ша/ТЕ/ТЖ/ФЖ</i>	43	41,81240
990	<i>у/ы/нша/ТЕ/ТЖ/ФЖ</i>	728	41,82774
----	-----	-----	-----
1028	<i>ма/ғандықтан/БЖ/К</i>	148	44,30215
1029	<i>па/ғандықтан/БЖ/К</i>	6	44,30228
1030	<i>ма/йтындықтан/БЖ/К</i>	10	44,30249
1031	<i>ме/гендіктен/БЖ/К</i>	68	44,30392
----	-----	-----	-----
1097	<i>сын/БР</i>	26530	51,59952
1098	<i>гін/БР</i>	1857	51,63865



Рет саны №	Түрленім форма	Абсолютті жиілік	Бір топ түрленім формалардың мәтінді қамту пайызы
1099	<i>йін/БР</i>	5848	51,76187
1100	<i>сін/БР</i>	15798	52,09476
-----	-----	-----	-----
1211	<i>на/йтын/дыз/ым/а/БЖ/Е/ СУБ/ТЖ/БС</i>	1	60,08739
1212	<i>ма/ған/дыз/ы/на/БЖ/Е/ СУБ/ТЖ/БС</i>	4	60,08747
-----	-----	-----	-----
1494	<i>ған/ымыз/да/Е/ТЖ/ЖС</i>	113	66,67895
2939	<i>ма/ды/БЖ/ЖӨШ</i>	3008	94,10825
2940	<i>на/ды/БЖ/ЖӨШ</i>	720	94,12342
2941	<i>бе/ді/БЖ/ЖӨШ</i>	260	94,12890
2942	<i>ме/ді/БЖ/ЖӨШ</i>	916	94,14820
2943	<i>не/ді/БЖ/ЖӨШ</i>	405	94,15674
2944	<i>ды/ЖӨШ</i>	159384	97,51524
2945	<i>ді/ЖӨШ</i>	117919	100
<b>Қосынды абсолютті жиілік – 4745688 – мәтін көлемі</b>	<b>Мәтінді қамту %</b>		

Кері-әліпбилі жиілік сөздікте түрленім формалар реті оларға қойылған белгіленімдердің соңғы әрпінің әліпби реті негізге алынады. Жоғарыда, мәселен, СУБ белгіленімімен басталып тұр. Себебі, белгіленімдер ішінен соңынан санағанда Б әрпі әліпби реті бойынша белгіленімдердің басы болып отыр, одан кейін ҚАРҒ деген белгіленім, себебі, соңынан санағандағы әліпби ретімен келесі әріп – Ғ әрпі. Әрі қарай осындай ретпен кері-әліпбилі жиілік сөздік жалғаса береді.

27-кестеде түрленім формалардың жиілікті әліпбилі сөздігінен үзінді берілген. Бұл сөздікте түрленім формалар мәтіндегі жиі кездесуіне қарай, яғни абсолютті жиілігі бойынша беріледі.

27-кесте. Қазақ тілінің түрленім формаларының жиілікті әліпбилі жиілік сөздігінен үзінді

Рет саны №	Түрленім форма	Абсолютті жиілік	Бір топ түрленім формалардың мәтінді қамту пайызы
1	<i>і/ТЖ</i>	335313	7,06564
2	<i>ы/ТЖ</i>	160419	10,44595
3	<i>ды/ЖӨШ</i>	159384	13,80445
4	<i>н/К</i>	135028	16,64972
5	<i>ты/ТС</i>	128558	19,35867
6	<i>ді/ЖӨШ</i>	117919	21,84343
7	<i>ған/Е</i>	115647	24,28032
8	<i>і/н/ТЖ/ТС</i>	108606	26,56883
9	<i>у/ТЕ</i>	105055	28,78253
10	<i>ын/К</i>	93445	30,75158
11	<i>ы/н/ТЖ/ТС</i>	88818	32,62313
12	<i>ген/Е</i>	70834	34,11573
13	<i>ге/БС</i>	70712	35,60575
14	<i>ға/БС</i>	65014	36,97571
15	<i>ң/ТЖ</i>	64458	38,33396
16	<i>да/ЖС</i>	59602	39,58988
17	<i>а/К</i>	55372	40,75666
-----	-----	-----	-----
25	<i>ның/ІС</i>	42065	48,67606
26	<i>ы/на/ТЖ/БС</i>	40097	49,52098
27	<i>ін/К</i>	38330	50,32866
-----	-----	-----	-----
206	<i>р/ы/ның/Е/ТЖ/ІС</i>	3039	90,03824
207	<i>шы/ӨФ</i>	3015	90,10177
208	<i>ма/ды/БЖ/ЖӨШ</i>	3008	90,16516

Рет саны №	Түрленім форма	Абсолютті жиілік	Бір топ түрленім формалардың мәтінді қамту пайызы
-----	-----	-----	-----
332	<i>сі/мен/ТЖ/КС</i>	1146	95,02207
333	<i>п/мі/К/ЖЖ</i>	1138	95,04605
334	<i>атын/ы/Е/ТЖ</i>	1135	95,06997
-----	-----	-----	-----
861	<i>ер/інде/Е/ТЖ/ЖС</i>	99	99,27511
862	<i>лер/іміз/дің/КЖ/ТЖ/ІС</i>	99	99,27720
863	<i>ма/йық/БЖ/БР</i>	99	99,27928
-----	-----	-----	-----
1676	<i>атын/дыз/ы/ның/Е/СУБ/ТЖ/ІС</i>	10	99,90958
1677	<i>атын/сың/Е/ЖЖ</i>	10	99,90979
1678	<i>бақ/ңыз/Е/ЖЖ</i>	10	99,91000
-----	-----	-----	-----
1720	<i>атын/сыз/Е/ЖЖ</i>	9	99,91881
1721	<i>ба/са/ңыз/БЖ/ШР/ЖЖ</i>	9	99,91900
1722	<i>бе/п/нін/БЖ/К/ЖЖ</i>	9	99,91919
-----	-----	-----	-----
1768	<i>ар/ымыз/да/Е/ТЖ/ЖС</i>	8	99,92789
1769	<i>атын/ымыз/ға/Е/ТЖ/БС</i>	8	99,92806
1770	<i>атын/ыңыз/Е/ТЖ</i>	8	99,92823
-----	-----	-----	-----
1823	<i>ар/ым/нан/Е/ТЖ/ШС</i>	7	99,93714
1824	<i>ар/ым/ның/Е/ТЖ/ІС</i>	7	99,93729
1825	<i>ар/ымыз/ға/Е/ТЖ/БС</i>	7	99,93744
-----	-----	-----	-----
1873	<i>тер/ің/нің/КЖ/ТЖ/ІС</i>	7	99,94452
1874	<i>у/ымыз/бен/ТЕ/ТЖ/КС</i>	7	99,94467



Жоғарыдағы түрленім формалардың жиілік сөздіктерінде түрленім форма саны – 2945. Абсолютті жиілігі ең жоғары түрленім формалар тәуелдік жалғаулары болып тұр. Мәселен -і/ТЖ 6 миллион сөзқолданыстан тұратын мәтінде 335313 рет кездесе, жуан варианты -ы/ТЖ 160419 рет кездесіп, екеуі қосылып, бүкіл мәтіннің 10 пайызын құрайды. Бұл жерде түрленім форманың жиілік сөздігі болғандықтан байқайтынымыз – тәуелді жалғаудың 3 жағының жеке-дара қолданысының жиі кездесуі. 27-кестедегі жиілігі жоғары реестрлік қатарда көпшілігі дара түрленім формалар болып тұр. Бұлардың ішінде 3-жақ тәуелдік жалғауының табыс септігімен қабаттаса қолданылуы ғана жоғары жиілікті көрсетеді. Мысалы, 8 және 11 реестрлік қатарда тұрған і/н/ТЖ/ТС мен ы/н/ТЖ/ТС жиілік сөздіктің жоғарғы жағында абсолютті жиіліктері жоғары болып тұр.

Келесі сөз болатын статистикалық мәлімет, ол **28-кестеде** келтірілген қазақ тілінің түрленім форма бірліктерінің кері-әліпбилі жиілік сөздігінің реестр «бойын» қамту пайызы жайлы болмақ. Кесте 4 бағанадан тұрады және әр бағанадағы мәліметтер атаулары кесте басында айтылған. 3-бағанадағы мәліметтерде жиілік сөздік принципі ұстаным етіліп, сөздік реестріндегі түрленім формалардың соңғы орында тұрған белгіленімдері негізге алынып, олардың реестр бойындағы кездесу саны (3-бағана) және реестр бойын қамту пайызы (4-бағана) қатаң кему тәртібімен орналастырылды. Демек, шартты белгіленім қойылған түрленім қосымшалар атауларының жиілік сөздіктегі белгіленім қойылған реестрлік түрленім формалардың ішінде неше рет кездесуіне қарай, көп кездескенінен бастап тізім құрайды.

**28-кесте. Қазақ тілінің түрленім форма бірліктерінің кері-әліпбилі жиілік сөздігінің реестр бойын қамту пайызы**

Рет саны	Түрленім форманың атауы	Түрленім форманың реестрдегі саны	Түрленім форманың реестр бойын қамту пайызы
1	2	3	4
1.	Жіктік жалғау (ЖЖ)	538	18,27%
2.	Табыс септік (ТС)	334	11,34%
3.	Тәуелдік жалғау (ТЖ)	318	10,80%
4.	Барыс септік (БС)	308	10,46%
5.	Ілік септік (ІС)	235	7,98%
6.	Көмектес септік (КС)	220	7,77%
7.	Шығыс септік (ШС)	218	7,40%
8.	Жатыс септік (ЖС)	196	6,66%



3-топқа: түрленім форманың реестрдегі саны – төмен: 1-ден көп, ал реестр бойын қамту пайызы – 0,05 % ÷ 0,30 % аралығы.

Бұл аталғандар **28-кесте** мәліметтері бойынша, 1-топқа жататындардың ішінде реестр бойында сан жағынан ең бірінші орынға ие болып тұрған түрленім формасы – **жіктік жалғау (ЖЖ)**. Аталған **жіктік жалғау** түрленім формасы реестрде 538 рет кездесіп, 2945 реестрлік бірліктің **18,27** пайызын қамтып жатқанына көз жеткізуге болады. Сонымен бірге осы 1-топтағы 9 түрленім формасы кері-әліпбилі жиілік сөздіктің реестр бойында сан жағынан 2516-ға тең реестрлік бірлік құрап, сөздіктегі барлық 2945 реестрлік бірліктің 85,74 пайызын қамтып жатқанын айтуға болады. Сол сияқты, 1-топтың ең көп реестрлік бірлік құрайтын **жіктік жалғаудан** басқа да реестр бойында сан жағынан жоғары сатыдағы қалған түрленім формаларына жататындар: **табыс септік, тәуелдік жалғау, барыс септік, ілік септік, көмектес септік, шығыс септік, жатыс септік, салыстыру форма** қосымшалары екеніне **28-кесте** мәліметтері бойынша көз жеткізе аламыз.

Сол сияқты 2-ші және 3-топтағы түрленім формаларының статистикалық сипаттамасына қысқаша тоқталайық.

Жоғарыда айтылғандай, 2-топтағы түрленім формалардың реестр бойындағы саны орта сатыға жатып, реестр бойында 390 реестрлік бірлікке ие екендігі және ол барлық реестр бойының 13,24 пайызын қамтитыны **28-кестедегі** мәліметтерден белгілі болып отыр. 2-топтың түрленім формалардың реестр бойында сан жағынан орташа сатыдағы түрленім формаларына жататындар: **өтіну форма, бұйрық рай, қатыстық форма, есімше, көсемше, көптік жалғау, тәуелдік форма, жоқтық форма, субстантив форма, шырай** сияқты 11 түрлі қосымша екендігі **28-кестедегі** статистикалық мәліметтерден белгілі болды.

Енді, соңғы 3-топтағы түрленім формаларына келетін болсақ, бұл топтағы 7 түрлі түрленім формалар реестр бойында сан жағынан 39 түрлі бірлік құрап, сөздіктегі барлық реестрлік бірліктің бар болғаны 1,34 пайызын қамтиды, яғни бұл 3-топтың реестрге қатынасы сан жағынан төмен деңгейде екендігінің айғағы деуге болады. 3-топтағы түрленім формалардың реестр бойында сан жағынан төменгі сатыдағы түрленім формаларына жататындар 7-еу: **қарғыс форма, тұйық етістік, шартты рай, жедел өткен шақ, болымсыздық жұрнақ, атрибут анықтауыш форма, қалау рай** (**28-кестедегі** статистикалық мәліметтерде).

Енді **29-кестеде** қазақ тілінің түрленім форма бірліктерінің кері-әліпбилі жиілік сөздігі бойынша құрастырылған түрленім форма құрамындағы әр түрленім қосымшаның және олардың абсолютті жиіліктері бойынша

біріктірілген топтарының мәтін көлемін қамту пайызының статистикалық мәліметтерін ұсынамыз.

**29-кесте. Қазақ тілінің әр түрленім формасының және олардың топтарының мәтін көлемін қамту пайызы**

Рет саны	Түрленім форманың атауы (белгіленім тізбегінің соңғысы)	Түрленім форманың реестрдегі саны	Түрленім форманың мәтін көлемін қамту пайызы (кему ретімен)	Бір топ түрленім форманың мәтін көлемін қамту пайызы
1.	Тәуелдік жалғау (ТЖ)	318	18,80%	18,80%
2.	Табыс септік (ТС)	334	12,23%	31,03%
3.	Жіктік жалғау (ЖЖ)	538	9,48%	40,51%
4.	Барыс септік (БС)	308	8,83%	49,34%
5.	Көсемше (К)	37	8,83%	58,17%
6.	Есімше (Е)	41	8,29%	66,46%
7.	Жатыс септік (ЖС)	196	6,48%	72,94%
8.	Жедел өткен шак (ЖӨШ)	7	5,96%	78,90%
9.	Шығыс септік (ШС)	218	3,53%	82,43%
10.	Көптік жалғау (КЖ)	31	2,30%	84,73%
11.	Ілік септік (ІС)	235	2,26%	86,99%
12.	Тұйық етістік (ТЕ)	7	2,25%	89,24%
13.	Көмектес септік (КС)	220	1,64%	90,88%
14.	Қатыстық форма (ҚФ)	49	1,44%	92,32%
15.	Салыстыру форма (СФ)	149	1,43%	93,75%
16.	Бұйрық рай (БР)	50	1,39%	95,14%



Рет саны	Түрленім форманың атауы (белгіленім тізбегінің соңғысы)	Түрленім форманың реестрдегі саны	Түрленім форманың мәтін көлемін қамту пайызы (кему ретімен)	Бір топ түрленім форманың мәтін көлемін қамту пайызы
17.	Шартты рай ( <b>ШР</b> )	7	1,36%	96,50%
18.	Болымсыздық жұрнақ ( <b>БЖ</b> )	5	1,31%	97,81%
19.	Функционалды жұрнақ ( <b>ФЖ</b> )	59	1,28%	99,09%
20.	Өтіну форма ( <b>ӨФ</b> )	58	0,25%	99,34%
21.	Шырай ( <b>Ш</b> )	10	0,15%	99,49%
22.	Тәуелдік форма ( <b>ТФ</b> )	27	0,23%	99,72%
23.	Атрибут анықтауыш форма ( <b>АТР</b> )	4	0,20%	99,92%
24.	Жоқтық форма ( <b>ЖФ</b> )	17	0,05%	99,97%
25.	Субстантив форма ( <b>СУБ</b> )	11	0,01%	99,98%
26.	Қарғыс форма ( <b>ҚАРҒ</b> )	7	0,01%	99,99%
27.	Қалау рай ( <b>ҚР</b> )	2	0,01%	100,00%
<b>Қосындысы:</b>		<b>2945</b>	<b>100%</b>	

**29-кестеде** қазақ тілінің түрленім форма бірліктерінің мәтін көлемін қамту пайызы, жиілік сөздік ұстанымындағыдай, кему тәртібімен орналас-ты. Олардың мәтінді қамту пайызы негізге алынып, **29-кестедегі** статисти-калық мәліметтерді **3 топқа** бөліп қарастырдық.

Сонымен, **1-топқа** мәтінді қамту пайызы жағынан 1-орынға ие 9 түрлі келесі түрленім формалар жатты: *тәуелдік жалғау, табыс септік, жіктік жалғау, барыс септік, көсемше, есімше, жатыс септік, жедел өткен шақ, шығыс септік*. Аталған түрленім формалар мәтін ішінде қолданыла келіп, барлық мәтіннің **82,43** пайызын қамтиды. Сол сияқты қолданыс жағынан **1-топтағы** түрленім формалардан төмен тұратын **2-топқа** жататын: *көптік*

*жалғау, ілік септік, тұйық етістік, көмектес септік, қатыстық форма, салыстыру форма, бұйрық рай, шартты рай, болымсыздық жұрнақ, функционалды жұрнақ* 10 түрленім форманың мәтін ішіндегі қолданысы барлық мәтіннің **16,66** пайызын қамтиды екен. Ал 1-топ пен 2-топтың түрленім формалары бірлесе қолданыла келіп, барлық мәтіннің **99,09** пайызын қамтитынын **29-кестеден** байқауға болады. Енді осы айтылған сипаттама-ларға ұқсас 3-топтың да статистикалық сипаттамасын сөз етелік.

**29-кестенің** мәліметтері бойынша, 3-топқа жататын 8 түрленім форма мыналар: *өтіну форма, шырай, тәуелдік форма, атрибут анықтауыш форма, жоқтық форма, субстантив форма, қарғыс форма, қалау рай*. Қайталана қолданылу негізінде аталған 8 түрленім форма барлық мәтіннің бар болғаны **0,91** пайызын ғана қамтиды, яғни 3-топтың мәтінді қамту пайызы жағынан **1 %**-да жетпейді деген сөз. Егер мәтіннің **82,43** пайызын қамтитын 1-топпен салыстыратын болсақ, 2-ші және 3-топтардағы түрленім формалардың бірлесе қолданысының өзі барлық мәтіннің **17,57** пайызын ғана қамтитынын байқауға болады.

**30-кесте** мәліметтері де жоғарыда аталған мәселелердің жалғасы деуге болады. Себебі, сөз болып жатқан статистикалық сипаттамалар сөздік пен мәтін арасындағы арақатынасты ажырата білуге септігін тигізеді. Мәселен, қарастырып отырған реестрлік бірліктің түрлері аз болып, реестр бойын қамту пайызы аз болса да, ол мәтін ішінде жиі қолданылудың нәтижесінде оның мәтінді қамту пайызы жоғары болуы мүмкін немесе керісінше, яғни реестр бойын қамту пайызы жоғары болса да, мәтін көлемін қамту пайызы аз болуы да мүмкін (сирек қолданыстағы бірліктер). Мысалы, **30-кестедегі** түрленім формалардың реестр бойы мен мәтін көлемін қамту пайыздары жайлы ақпарат беретін 4-ші және 5-бағаналарда көрініс тапқан 1-ші және 3-топтарға қатысты пайыздық көрсеткіштер осы айтылғанға айғақ бола алады. Осындай мәселелер жайлы дұрыс қорытынды жасауға мән берілуі қажет деп білеміз.

Енді **30-кестедегі** мәліметтерге келетін болақ, кесте 5 бағанадан тұрады: **1-бағанада** – топтардың рет саны; **2-бағанада** – топтардың рет сан аралығы; **3-бағанада** – абсолютті жиіліктер аралығы; **4-бағанада** – түрленім формалардың реестр бойын қамту пайызы; **5-бағанада** – түрленім формалардың мәтін көлемін қамту пайызы. Кесте мәліметтеріне зер салып қарасаңыздар, бірдей реестрлік аралықтарда орналасқан түрленім формалардың реестр бойы мен мәтін көлемін қамту пайыздары 1-ші және 3-топтарда айтарлықтай айырым табатынын байқауға болады, ал мұндай жағдай 2-реттік топ үшін неге бір-бірімен шамалас дәрежеде? Әрине, бұл жағдай, негізінен, жиілік сөздіктегі реестрлік бірліктердің абсолютті жиіліктерінің қатаң түрде кему тәртібімен орналасуында деуге болады. Бірақ сирек жиілікті бірліктердің

сан жағынан ұлғаюының себебі неде? Бұл мәселе барлық тілдерге ортақ заңдылық па? – деген сауалдар да туындауы мүмкін.

**30-кесте. Қазақ тілінің түрленім формалар топтарының жиілік сөздіктің реестр бойы мен мәтін көлемін қамту пайызы**

Топ реті	Топтардың рет сан аралығы	Абсолютті жиіліктер аралығы	Түрленім формалардың реестр бойын қамту пайызы	Түрленім формалардың мәтін көлемін қамту пайызы
1	2	3	4	5
1.	$1 \leq n_i \leq 87$	$10000 \leq F_i \leq 400000$	2,954 %	76,580 %
2.	$88 \leq n_i \leq 860$	$100 \leq F_i \leq 10000$	26,248 %	22,693 %
3.	$861 \leq n_i \leq 2945$	$1 \leq F_i \leq 99$	70,798 %	0,790 %
		Қосындысы:	100 %	100 %

**31-кесте Қазақ тілінің түрленім формаларының жоғары жиілікті тобы**

Рет саны	Түрленім форма	Абсолютті жиілік	Бір топ түрленім формалардың мәтінді қамту пайызы
1	<i>i/ТЖ</i>	335313	7,06564
2	<i>ы/ТЖ</i>	160419	10,44595
3	<i>ды/ЖӨШ</i>	159384	13,80445
4	<i>n/К</i>	135028	16,64972
5	<i>ты/ТС</i>	128558	19,35867
6	<i>di/ЖӨШ</i>	117919	21,84343
7	<i>ған/Е</i>	115647	24,28032
8	<i>i/n/ТЖ/ТС</i>	108606	26,56883
9	<i>у/ТЕ</i>	105055	28,78253
10	<i>ын/К</i>	93445	30,75158

31-кестеде қазақ тілінің түрленім формаларының жоғары жиілікті тобының сан жағынан 10-ға тең және мәтін көлемінің 31 пайызын қамтып жатқан бөлігін беріп отырмыз.

Кестеден көріп отырғанымыздай, түрленім формалардағы жиілік негізінен жалаң қосымшалар жиілігімен шамалас болып тұр. Бұдан қазақ тілінде түрленім формалары дара, жалаң тұлғада жиірек кездесетінін байқаймыз.

Сонымен қатар, қосымшалардың түрленім формалардың жиілік сөздігіндегі реестр бойында кездесуіне қарай сан жағынан ең бірінші орынға ие болып тұрған түрленім формасы – **жіктік жалғау (ЖЖ)** реестрде 538 рет кездесіп, 2945 реестрлік бірліктің **18,27** пайызын қамтиды. Бұдан қазақ тілінде сөздер жіктік жалғауымен жиі түрленіп қолданылатынын көреміз, яғни бұл жіктік жалғауының предикаттық қызметімен тікелей байланысты болса керек.

Сонымен бірге **жіктік жалғаудан** басқа да реестр бойында сан жағынан жоғары сатыдағы түрленім формаларына жататындар: **табыс септік, тәуелдік жалғау, бағыс септік, ілік септік, көмектес септік, шығыс септік, жатыс септік, салыстыру форма** қосымшалары екені анықталды. Бұл қазақ тіліндегі септік жалғауларының сөз байланыстырушылық қызметімен байланысты. Ал тәуелдік жалғаудың реестрдегі түрленім формалар құрамында көп кездесуі оның жалпы жиілік сөздіктегі бірінші орынға шығуына да ықпал еткен. Ал салыстыру формасының реестр бойындағы жоғары жиіліктегі топқа енуі *-дай/дей* қосымшасының кез келген сөзге жалғанып, сөз бойында жалғаулар сияқты еркін қолданысымен байланысты.

29-кестедегі статистикалық мәліметтер бойынша тұжырым жасайтын болсақ, жіктік жалғау түрленім форма реестрінде жиі кездесуі жағынан, яғни реестрді қамту пайызы жағынан ең жоғары жиілікті көрсетіп, тізімде бірінші болып тұрған болса, мәтінді қамту пайызы жағынан жоғары жиілікті тәуелдік жалғауы алып кетеді. Тәуелдік жалғаулары 6 миллион сөзқолданыстан тұратын мәтіннің 18,80 пайызын құрайды екен. Ал табыс септігі 12, 23 пайызын, жіктік жалғау – 9,48 пайызын, бағыс септігі – 8,83 пайызын қамтиды. Осы төрт түрленім қосымша қосылып, мәтіннің 50 пайызға жуығын қамтиды екен.

## 4 ҚАЗАҚ ТІЛІНДЕГІ СӨЗДЕРДІҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ

### 4.1 Қазақ тіліндегі сөздердің жиілік сөздігін құрастырудың қажеттілік негіздемесі

Қазіргі кезде қазақ лексикологиясының аса бір маңызды мәселесінің бірі – қазіргі қазақ тілінің қолданыстағы лексикалық жүйесінің шекарасын айқындау. «Қазақ тілінің жиілік сөздігі», ең алдымен, жиі қолданыстағы сөздік қорымыздың шекарасын анықтап алу үшін өте қажет. Жиілік сөздіктерді құрастыру ісі, басқа да лексикографиялық еңбектер сияқты, әр кезеңде де әртүрлі теориялық және қолданбалы бағыттардағы зерттеу жұмыстарын жүргізу үшін маңызы зор өзекті тақырыптардың бірі болып келеді.

Мәтін мен оның жиілік сөздігі бірліктерінің арақатынасына тоқтайтын болсақ, онда ең алдымен мына жайтқа зер салайық. Тілдік бірліктердің деңгейі жоғарылаған сайын олардың формалды түрде жігін айыру қиындай түсетіні белгілі. Мұндай жағдай тіліміздегі лексика құрамының мағыналық аспектісінің көпжақтылық сипатына байланысты болады. Әсіресе, бұл сөздік қор бірліктерінің ішкі және сыртқы тілдік факторларының әсерінен бір-бірімен өзара тығыз және күрделі түрде әрекеттесуінің нәтижесінен болуы мүмкін. Сөздердің тілімізде қолданыс табуы кездейсоқтық оқиға деп есептесек, оның заңды не заңсыз құбылыс екендігі тек ықтималдықтар теориясы мен математикалық статистика әдістері арқылы ғана анықталады. Тек осы ғылымдар саласы ғана қайталанатын және біркелкілік сипаттағы оқиғалар аясымен шұғылдана алады. Сөздік қор мәселесін зерттеу барысында тілге әсер ететін қалам иесінің (автордың) әлемге деген көзқарасының барлық жақтары – әлеуметтік, тұрмыстық жағдай ерекшеліктері және т.б. бірдей ескерілуі тиіс. Себебі олар лексикалық қорға өз танбасын қалдырумен бірге, тілді статистикалық жолмен зерттеудің қол жеткізер нысанына айналады.

Орыс тіл білімінде бір тілдің барлық стильдерін қамтыған жиілік сөздігі жарық көрген. Бұл сөздіктің орыс тілін дамытуда, орыс тілін өзге ұлт өкілдеріне үйретуде маңызы зор болды.

Ана тіліміздің немесе басқа да туыстас (не туыстас емес) тілдердің әр стильдік мәтіндеріне қатысты жасалған жиілік сөздіктер сол тілдің лексикалық құрылымын терең зерттеуде, лексиканың жалпы функционалдық стильге ортақ бөлігін ажыратуда септігі мол. Сондай-ақ, жалпы халықтық лексиканың, кірме сөздер мен терминдердің, неологизмдердің және басқа лексикалық топтардың тілдің жалпы лексика жүйесінен алатын орны жөнінде де жиілік сөздік бойынша көптеген сандық және сапалық мәліметтер алуға



СӨЗ/сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
е/ет.	26122	6,6756
өз/ес.	23635	7,7997
кел/ет.	22566	8,8730
ал/ет.	22522	9,9442
ол/ес.	22262	11,0031
да/шл.	22013	12,0501
де/шл.	20282	13,0147
сол/ес.	17723	13,8577
мен/ес.	17341	14,6825
айт/ет.	16410	15,4630
бір/са.	15276	16,1895
бер/ет.	14691	16,8883
отыр/ет.	12382	17,4772
осы/ес.	11818	18,0393
жөк/мд.	11415	18,5822
сөз/зт.	11309	19,1201
тұр/ет.	11022	19,6443
кет/ет.	10635	20,1501
жүр/ет.	10555	20,6522
бір/ес.	10344	21,1442

Көріп отырғанымыздай, *бол* және *де* етістіктерінің жиілігі өте жоғары. Жиілік бойынша жоғары қолданысты есімдіктер, көмекші етістіктер, шылаулар құрайды.

Бүгінгі таңда мемлекеттік тіліміздің қолданыс қызметін арттыру мақсатындағы заманауи қажеттіліктерге сәйкес қолданбалы бағыттағы ғылыми жобаларға ерекше мән беріліп отырғандығы белгілі. Бұл ретте 2016 жылы Білім және ғылым министрі Сағадиевтің тапсыруымен мемлекеттік тілді оқытудың тиімді әдістемесін жасау мақсатында лексика-грамматикалық минимумдар үшін қажетті қазақ тілінің жиілік сөздігін құрастыру ісі А.Байтұрсынұлы атындағы Тіл білімі институтына тапсырылған болатын. Осы тапсырма бойынша институт қысқа мерзімнің ішінде «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» жарық көрді [8.]. «Жалпы білім берудегі қазақ тілінің жиілік сөздігін» құрастыруға көркем әдебиет, публицистикалық стиль, ғылыми, публицистикалық, сөйлеу стильдер мәтіндері қамтылған, яғни қазақ тілінің 5 функционалдық стилі сөздіктің базасын құрайды. 36 265 сөзтізбеден (сөз, лексикалық бірлік) тұратын бұл сөздік дәстүрлі сөздіктер





СӨЗ/сөз табы	Барлығы
қандай/ес	21378
екен/ет	21328
біз/ес	21267
сол/ес	20532

Бұл сөздікте де *бол* көмекші етістігінің жиілігі басқаларына қарағанда өте жоғары. Бұл кестеден де көмекші етістіктер, шылаулар мен есімдіктердің жоғары қолданыстағы сөздер екендігін байқаймыз. Сонымен қатар жоғары-да берілген біріктірілген жиілік сөздіктегі көрсеткіштермен ұқсас.

Біраз уақыттан бері тоқтап қалған жиілік сөздіктер құрастыру ісі министрдің арнайы тапсырысымен шұғыл түрде қолға алынуы мемлекеттік тіліміздің қолданыс аясын дамытудағы келелі істердің бірі болды деуге болады.

### 4.3 Қазақ тіліндегі сөздердің жиілік сөздігінің сипаттамасы

Жиілік сөздік жасаудың әдіс-тәсілдері әртүрлі. Соның бірі – мәтінді сөзформалар сөздігіне түсіріп, сөзді түбірге келтіру. Екіншісі – мәтінді морфологиялық анализатор арқылы өткізіп, түбір мен оған қойылатын сөз табы белгісімен бірге автоматты жолмен алу. Біріншісі көп қол жұмысын керек етеді. Мұндайда қосымша жалғанғанда түбір сөздердің омонимдерін ажырату мүмкін болса, түбір формадағы сөздерге келгенде олардың сөз табына қатысын анықтау қиын болады.

Екіншісі оңай, тез жүзеге асырылғанмен, жүз пайыздық шынайылықты бермейді. Бұл әдісте те омонимдер автоматты түрде ажыратылмайды. Автоматты жолмен омонимдері ажыратылған жиілік сөздіктерді тек омоним ажыратылған мәтіндерден ғана алуға болады.

Осы орайда Қолданбалы лингвистика бөлімі қызметкерлері проза жанры (М.Әуезов, Ә.Кекілбаев т.б. ) бойынша 500 000 сөзқолданыстан тұратын мәтіндерге «қолдап» морфологиялық белгіленім қойған болатын. Қолмен қойылған морфологиялық белгіленімнен үзінді келтірсек:  
*34-кесте.*



Жиілік сөздіктердің көп тараған түрлері мынадай:

- 1) әліпби жиілік сөздік;
- 2) жиілік сөздік;
- 3) кері әліпби жиілік сөздік;

Әліпби жиілік сөздікте көрініс табатын ең бірінші (қатарда) – сөздік бірлігінің реттік сандары, екінші – мәтіннен бөлініп алынған әртүрлі сөздің (сөзтұлғаның) қатаң әліпби тәртібімен орналасқан тұлғасы және үшінші – сөздің абсолютті жиілігі.

Жиілік сөздіктердің толық түрінде осы аталғандардан басқа: «жиынтық-абсолютті жиілік», «қатынастық жиілік», «жиынтық қатынастық жиілік» және сонымен бірге **бір топ сөздердің мәтінді қамтуының пайыздық шамалары** беріледі. Жиілік сөздіктер жасау (түзу) арнайы компьютерлік бағдарламалар көмегімен іске асады.

Әліпби жиілік сөздік зерттеушіге қажет деген сөздің (оның туынды тұлғаларының да) қолданылу дәрежесін, яғни оның жиілігін табуға, зерттеуге алынған мәтіндер лексикасының әліпбилік құрамының статистикасын анықтау үшін аса маңызды. Сондықтан оны тілдің әліпбилік тұрғыдағы статистикалық үлгісі (моделі) деп те атауға болады.

Тіл мәдениетін зерттеуші ғалымдар мұндай әліпби жиілік сөздіктен бағалы мәліметтер таба алады. Сол сияқты, әр стильдік мәтіндердің өзіндік тіл ерекшеліктерін анықтауға және жалпы тіл біліміне қатысты көптеген теориялық мәселелердің басын ашуға мүмкіндік туады деуге болады.

Төменде осы аталған әліпбилі жиілік сөздіктен үзіндіні **35-кесте** түрінде ұсынып отырмыз:

**35-кесте. Қазақ тіліндегі сөздің әліпбилі жиілік сөздігі (үзінді)**

№ р/с	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
1	<i>а/од.</i>	1	0,003
2	<i>абажжа/зт.</i>	2	0,010
3	<i>абажжадай/сн.</i>	1	0,014
4	<i>абжжылан/зт.</i>	1	0,017
5	<i>абзал/мд.</i>	2	0,024
6	<i>абырой/зт.</i>	1	0,028
7	<i>абысын/зт.</i>	2	0,035
8	<i>автор/зт.</i>	1	0,038

№ р/с	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
9	<i>авторлық/зт.</i>	1	0,042
10	<i>ага/зт.</i>	10	0,077
11	<i>ағайын-туған/зт.</i>	1	0,080
12	<i>ағайын/зт.</i>	6	0,101
13	<i>ағар/ет.</i>	2	0,108
14	<i>ағаш/зт.</i>	8	0,136
15	<i>ағзам/зт.</i>	2	0,143
16	<i>ағыл-тегіл/сн.</i>	1	0,146
17	<i>ағында/ет.</i>	1	0,150
18	<i>адал/сн.</i>	3	0,160
19	<i>адам/зт.</i>	64	0,383
20	<i>адамдық/зт.</i>	2	0,390
21	<i>адас/ет.</i>	5	0,408
22	<i>адыр-адыр/зт.</i>	1	0,411
23	<i>адыр/зт.</i>	7	0,435
24	<i>адырай/ет.</i>	1	0,439
25	<i>ажал/зт.</i>	1	0,442
26	<i>ажар/зт.</i>	8	0,470
27	<i>ажарлы/сн.</i>	2	0,477
28	<i>ажарсыз/сн.</i>	1	0,481
29	<i>ажым/зт.</i>	1	0,484
30	<i>ажырай/ет.</i>	1	0,488
31	<i>ажырат/ет.</i>	1	0,491
32	<i>аз-аздап/үс.</i>	5	0,509
33	<i>аз/сн.</i>	59	0,714
34	<i>аз/үс.</i>	2	0,721
35	<i>азай/ет.</i>	5	0,738
36	<i>азамат/зт.</i>	3	0,749
37	<i>азан-қазан/үс.</i>	1	0,752
38	<i>азгана/үс.</i>	6	0,773

№ р/с	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
39	<i>азгантай/сн.</i>	2	0,780
40	<i>азгыр/ет.</i>	1	0,784
41	<i>аздық/зт.</i>	1	0,787
42	<i>азық/зт.</i>	8	0,815
43	<i>азын-аулақ/сн.</i>	2	0,822
44	<i>азын-аулық/ұс.</i>	1	0,825
45	<i>азына/ет.</i>	3	0,836
46	<i>ай-ай/од.</i>	1	0,839
47	<i>ай-жұлдыз/зт.</i>	1	0,843
48	<i>ай/зт.</i>	24	0,927
49	<i>ай/од.</i>	1	0,930
50	<i>ай/шл.</i>	8	0,958
51	<i>айгай/зт.</i>	5	0,975
52	<i>айгайла/ет.</i>	1	0,979
53	<i>айзақ/зт.</i>	1	0,982
54	<i>айғыз-айғыз/сн.</i>	1	0,986
55	<i>айда/ет.</i>	3	0,996
56	<i>айдала/зт.</i>	2	1,003
57	<i>айдау/зт.</i>	1	1,007
58	<i>айдаһар/зт.</i>	1	1,010
59	<i>айдаһарша/ұс.</i>	1	1,014
60	<i>айдынды/сн.</i>	1	1,017
61	<i>айқайла/ет.</i>	1	1,021
62	<i>айқара/ұс.</i>	1	1,024
63	<i>айқын/сн.</i>	6	1,045
64	<i>айқында/ет.</i>	1	1,048
65	<i>айқыш-үйқыш/сн.</i>	1	1,052
66	<i>айла-тәсіл/зт.</i>	1	1,055
67	<i>айла/зт.</i>	6	1,076
68	<i>айлапат/сн.</i>	3	1,087
----	-----	-----	-----

Енді «Әліпбилі жиілік сөздіктен» қазақ тілі мәтінінің статистикалық сипаттамасын жазу үшін зерттеушіге қажетті келесі 36 мен 37-кестелерде көрініс тапқан статистикалық мәліметтерді сөз етелік.

**36-кесте. Әліпбилі жиілік сөздігіндегі бастапқы әріпке қатысты сөздердің саны мен реестр бойын қамту пайызы**

Рет саны (i=)	Әліпбилі жиілік сөздігіндегі сөздердің бастапқы әріптері	Бастапқы әріптен басталатын сөздердің саны (N=)	Бастапқы әріптен басталатын сөздердің реестр бойын қамту пайызы (%)
1.	<i>А (а)</i>	390	8,920 %
2.	<i>Ә (ә)</i>	124	2,066 %
3.	<i>Б (б)</i>	394	13,943 %
4.	<i>В (в)</i>	1	0,003 %
5.	<i>Г (г)</i>	11	0,066 %
6.	<i>Ғ (ғ)</i>	13	0,509 %
7.	<i>Д (д)</i>	118	4,939 %
8.	<i>Е (е)</i>	152	4,918 %
9.	<i>Ж (ж)</i>	430	9,930 %
10.	<i>З (з)</i>	37	0,293 %
11.	<i>И (и)</i>	35	0,338 %
12.	<i>К (к)</i>	331	10,014 %
13.	<i>Қ (қ)</i>	548	10,149 %
14.	<i>Л (л)</i>	17	0,129 %
15.	<i>М (м)</i>	195	3,16 %
16.	<i>Н (н)</i>	63	0,853 %
17.	<i>О (о)</i>	122	4,124 %
18.	<i>Ө (ө)</i>	82	2,212 %
19.	<i>П (п)</i>	53	0,609 %
20.	<i>Р (р)</i>	28	0,321 %
21.	<i>С (с)</i>	380	6,308 %
22.	<i>Т (т)</i>	471	8,018 %
23.	<i>У (у)</i>	18	0,32 %
24.	<i>Ұ (ұ)</i>	79	1,052 %
25.	<i>Ү (ү)</i>	82	2,167 %

Рет саны (i=)	Әліпбилі жиілік сөздігіндегі сөздердің бастапқы әріптері	Бастапқы әріптен басталатын сөздердің саны (N=)	Бастапқы әріптен басталатын сөздердің реестр бойын қамту пайызы (%)
26.	<b>Ф (ф)</b>	4	0,02 %
27.	<b>Х (х)</b>	14	0,279 %
28.	<b>Һ (һ)</b>	2	0,014 %
29.	<b>Ш (ш)</b>	226	2,762 %
30.	<b>Ы (ы)</b>	57	0,512 %
31.	<b>І (і)</b>	38	1,00 %
32.	<b>Э (э)</b>	1	0,003 %
33.	<b>Я (я)</b>	5	0,049 %
		$\Sigma=4521$	$\Sigma=100\%$

**36-кестедегі** статистикалық мәліметтер **37-кестеде** де орын алған, бірақ ол кестедегі мәліметтер Әліпбилі жиілікті сөздік ұстанымен емес, жиілікті әліпбилі сөздіктің принципімен, яғни *абсолютті жиілік, қатынастық жиілік, қосынды жиілік, реестрді қамту пайызы, мәтінді қамту пайызы* шамалары қатаң кему тәртібімен орналасады.

**37-кесте. Бастапқы әріпке қатысты сөздердің жиілікті әліпбилі сөздігіндегі реестр мен мәтінді қамту статистикасы**

Рет саны (i=)	Жиілікті әліпбилі сөздіктің реестріндегі сөздердің бастапқы әріптері	Реестрдегі бастапқы әріпке қатысты сөздер саны	Бастапқы әріпке қатысты сөздердің реестрді қамту пайызы	Рет саны (i=)	Жиілікті әліпбилі сөздіктің реестріндегі сөздердің бастапқы әріптері	Бастапқы әріпке қатысты сөздердің қосынды абсолютті жиілігі	Бастапқы әріпке қатысты сөздердің мәтінді қамту пайызы
1	2	3	4	5	6	7	8
1.	<b>Қ (қ)</b>	548	12,121 %	1.	<b>Б (б)</b>	4003	13,943 %
2.	<b>Т (т)</b>	471	10,418 %	2.	<b>Қ (қ)</b>	2914	10,150 %
3.	<b>Ж (ж)</b>	430	9,511 %	3.	<b>К (к)</b>	2875	10,014 %
4.	<b>Б (б)</b>	394	8,715 %	4.	<b>Ж (ж)</b>	2851	9,930 %
5.	<b>А (а)</b>	390	8,626 %	5.	<b>А (а)</b>	2561	8,920 %
6.	<b>С (с)</b>	380	8,405 %	6.	<b>Т (т)</b>	2302	8,018 %

Рет саны (№)	Жиілікті әліпбилі сөздіктің реестріндегі сөздердің бастапқы әріптері	Реестрдегі бастапқы әріпке қатысты сөздер саны	Бастапқы әріпке қатысты сөздердің реестрді қамту пайызы	Рет саны (№)	Жиілікті әліпбилі сөздіктің реестріндегі сөздердің бастапқы әріптері	Бастапқы әріпке қатысты сөздердің қосынды абсолютті жиілігі	Бастапқы әріпке қатысты сөздердің мәтінді қамту пайызы
7.	<i>К (к)</i>	331	7,321 %	7.	<i>С (с)</i>	1811	6,308 %
8.	<i>Ш (ш)</i>	226	4,999 %	8.	<i>Д (д)</i>	1418	4,939 %
9.	<i>М (м)</i>	195	4,313 %	9.	<i>Е (е)</i>	1412	4,918 %
10.	<i>Е (е)</i>	152	3,362 %	10.	<i>О (о)</i>	1184	4,124 %
11.	<i>Ә (ә)</i>	124	2,743 %	11.	<i>М (м)</i>	907	3,159 %
12.	<i>О (о)</i>	122	2,698 %	12.	<i>Ш (ш)</i>	793	2,762 %
13.	<i>Д (д)</i>	118	2,610 %	13.	<i>Ө (ө)</i>	635	2,212 %
14.	<i>Ө (ө)</i>	82	1,814 %	14.	<i>Ү (ү)</i>	622	2,166 %
15.	<i>Ү (ү)</i>	82	1,814 %	15.	<i>Ә (ә)</i>	593	1,762 %
16.	<i>Ұ (ұ)</i>	79	1,747 %	16.	<i>Ұ (ұ)</i>	302	1,052 %
17.	<i>Н (н)</i>	63	1,393 %	17.	<i>І (і)</i>	287	0,999 %
18.	<i>Ы (ы)</i>	57	1,261 %	18.	<i>Н (н)</i>	245	0,853 %
19.	<i>П (п)</i>	53	1,172 %	19.	<i>П (п)</i>	175	0,609 %
20.	<i>І (і)</i>	38	0,840 %	20.	<i>Ы (ы)</i>	147	0,512 %
21.	<i>З (з)</i>	37	0,818 %	21.	<i>Ғ (ғ)</i>	146	0,508 %
22.	<i>И (и)</i>	35	0,774 %	22.	<i>И (и)</i>	97	0,338 %
23.	<i>Р (р)</i>	28	0,619 %	23.	<i>Р (р)</i>	92	0,320 %
24.	<i>У (у)</i>	18	0,398 %	24.	<i>У (у)</i>	92	0,320 %
25.	<i>Л (л)</i>	17	0,376 %	25.	<i>З (з)</i>	84	0,292 %
26.	<i>Х (х)</i>	14	0,309 %	26.	<i>Х (х)</i>	80	0,279 %
27.	<i>Ғ (ғ)</i>	13	0,287 %	27.	<i>Л (л)</i>	37	0,129 %
28.	<i>Г (г)</i>	11	0,243 %	28.	<i>Г (г)</i>	19	0,066 %
29.	<i>Я (я)</i>	5	0,110 %	29.	<i>Я (я)</i>	14	0,049 %
30.	<i>Ф (ф)</i>	4	0,088 %	30.	<i>Ф (ф)</i>	6	0,021 %
31.	<i>Һ (һ)</i>	2	0,044 %	31.	<i>Һ (һ)</i>	4	0,014 %
32.	<i>В (в)</i>	1	0,022 %	32.	<i>В (в)</i>	1	0,003 %
33.	<i>Ә (ә)</i>	1	0,022 %	33.	<i>Ә (ә)</i>	1	0,003 %
		$\Sigma=4521$	$\Sigma=100\%$			$\Sigma=28710$	$\Sigma=100\%$

37-кестедегі, яғни «Бастапқы әріпке қатысты сөздердің жиілікті әліпбилі сөздігіндегі реестр мен мәтінді қамту статистикасы» атты кестенің 3, 4, 7, 8 бағаналарындағы сандық мәліметтер қатаң түрде кему тәртібімен ор-



наласқан. Аталған сандық мәліметтер екі мәселені қамтып жатыр: *бірінші* – қазақ әліпбиінде орын алған әріптерден басталып жазылатын сөздер сөздік бойының (реестрдің) қанша пайызын қамтиды; *екінші* – бастапқы әріпке қатысты сөздердің қайталана қолданыла келіп, мәтін көлемінің қанша пайызын қамтитынына қатысты мәселелер. Біз ұсынып отырған үш түрлі жиілік сөздікте, яғни **Әліпбилі жиілік сөздік**, **жиілікті әліпбилі сөздік**, **кері-әліпбилі жиілік** сөздіктерді зерттеуші тілші ғалымдар алға қойған өз мақсаттары үшін пайдалана алады.

Енді **37-кестедегі** жиілікті әліпбилі сөздіктің реестрлік қатарындағы әрбір сөздің бастапқы әрпіне қатысты статистикалық мәліметтерге тоқталайық. Жоғарыда айтылғандай, бұл деректер жиілік сөздік түрінде, яғни сөздіктің реестріндегі ең жиі қолданыстағы сөзден басталып, оның жиілігінің біртіндеп кему тәртібімен орналасатыны айтылған болатын. Осыған орай, жиілік сөздік реестріндегі сөздер санын, реестр бойындағы сөздердің бастапқы әрпіне қарай, **ең көп кездесетін, орташа санмен кездесетін және ең аз кездесетін** деп үш топқа бөліп қарастырамыз.

Мәселен, Әліпбилі жиілік сөздіктің реестріндегі бастапқы әріпке қатысты сөздер санын таңдау мақсатымен біз 3 орын алатын сандарды **1-ші топқа**, кездесу саны 118-548 аралығындағыны, яғни ең көп сөздер санына қатысты. Осы жерде жеке айтуға тұрарлық мәселе, ол 1-ші және 2-ші реттік санымен орналасқан **1-ші топты** бастап тұрған **«Қ»** мен **«Т»** әріптері. Бұл екі әріптен басталатын реестрлік сөздердің сандық үлесі барлық реестрлік сөздер ұзындығының **22,539 пайызын** қамтып жатқан **1019 реестрлік** сөз.

Кездесу саны 2 орын ғана алатындарды 2-ші топқа, кездесу саны 11-82 аралығындағыны, яғни кездесу саны орташа және 3-ші топқа кездесу саны өте төмен 1-5 аралығындағыларды алуға ұйғардық. Сондықтан, **1-ші топқа** енетін кестедегі рет саны 1-13 аралығындағы – **Қ (қ), Т (т), Ж (ж), Б (б), А (а), С (с), К (к), Ш (ш), М (м), Е (е), Ә (ә), О (о), Д (д)** әріптерінен басталатын реестрлік сөздер. Осы әріптерден басталатын реестрлік сөздердің жалпы саны – **3881**-ге тең және жиілік сөздіктің реестрлік тізімнен алатын пайыздық үлесі – **85,84** пайыз. **2-ші топқа** енетін кестедегі рет саны 14-28 аралығындағы – **Ө (ө), Ү (ү), Ұ (ұ), Н (н), Ы (ы), І (і), 3 (з), И (и), Р (р), (у), Л (л), Х (х), F (з), Г (з)** әріптерінен басталатын реестрлік сөздер. Мұндай әріптерден басталатын реестрлік сөздердің жалпы саны – **627**-ге тең жиілік сөздіктің реестрлік тізімнен алатын пайыздық үлесі – **13,87** пайыз. Енді соңғы, **3-ші топқа** енетін кестедегі рет саны 29-33 аралығындағы – **Я (я), Ф (ф), Ы (һ), В (в), Э (э)** әріптерінен басталатын реестрлік сөздер. Бұл әріптерден басталатын реестрлік сөздердің жалпы

саны – 13-ке тең, ал жиілік сөздіктің реестрлік тізімнен алатын пайыздық үлесі – 0,29 пайыз.

Келесі қарастыратынымыз *37-кестедегі*, яғни «Бастапқы әріпке қатысты сөздердің жиілікті әліпбилі сөздігіндегі реестр мен мәтінді қамту статистикасы» атты кестенің 6-бағанасындағы сөздік реестріндегі сөздердің бастапқы әріптері және 7-бағанадағы бастапқы әріпке қатысты сөздердің қосынды абсолютті жиілігі, 8-бағанадағы бастапқы әріпке қатысты сөздердің мәтінді қамту пайызы жайындағы сандық мәліметтер, яғни реестрдегі сөздердің бастапқы әрпіне қатысты сөздердің қайталана қолданыла келіп, мәтін көлемінің қанша пайызын қамтитынына қатысты мәліметтер.

Аталған мәліметтерге көз жеткізу үшін біз алдыңғы әдіс бойынша, бастапқы әріпке қатысты сөздердің кему тәртібімен орналасқан қосынды абсолютті жиіліктерін үш топқа бөліп қарастырдық. Мәселен, Әліпбилі жиілік сөздіктің реестріндегі бастапқы әріпке қатысты сөздер санын таңдау мақсатымен біз 4 орын алатын сандарды *1-ші топқа* жатқыздық, қосынды абсолютті жиіліктері 1184-4003 аралығындағыны, яғни ең көп сөздер санына қатысты топ. Бұл топқа жататындар реестрдегі келесі 10 әріптен басталатын реестрлік сөздер: *Б (б), Қ (қ), К (к), Ж (ж), А (а), Т (т), С (с), Д (д), Е (е), О (о)*. Бірінші топқа жататын сөздердің қосынды абсолютті жиілігі *23331-ге* тең, ал ол сөздердің мәтінді қамту пайызы 81,26-ға тең. Сол сияқты, *2-топқа* рет сандары 11-21 аралығындағы, реестрлік сөздердің бастапқы әріптері: *М (м), Ш (ш), Ө (ө), Ү (ү), Ә (ә), Ұ (ұ), І (і), Н (н), П (п), Ы (ы), Ғ (ғ)* болатын, ал қосынды абсолютті жиіліктері 146-907 аралығындағы сандарға тең және мәтінді қамту пайызы – 16,59-ға тең орташа жиілікті топты жатқыздық.

Енді соңғы *3-топқа 37-кестедегі* рет сандары 22-33 аралығындағы мына 12 әріпті: *И (и), Р (р), У (у), З (з), Х (х), Л (л), Г (г), Я (я), Ф (ф), Һ (һ), В (в), Э (э)*. Олардың қосынды абсолютті жиіліктері *97-1* аралығында сандарға тең және мәтінді қамту пайызы – 1,83-ке тең ең төмен жиілікті топты жатқызуды ұйғардық.

Сонымен, 500 000 сөзқолданыстан тұратын мәтінде Қ және Т әрпінен басталатын сөз саны 1019 реестрлік сөзді құрап, барлық реестрлік сөздер ұзындығының *22,539 пайызын* қамтиды екен. 37-кестенің бірінші бағанасындағы 13 әріп мәтінде сөз басында жиі кездесетін әріптер болып саналады. Олар біріге келе 3881 сөзді құрап, реестрлік тізімнің 85,84 пайызын қамтиды. Бұл жиілік мәтіні көркем шығармалардан алынғандықтан, Я, Ф, В, В, Э әрпінен басталатын сөздер бірлі-жарым ғана кездескен.

37-кестенің 2-бағаны туралы статистикалық мәліметтерден мынаны ұғуға болады. Сөз басында кездесетін әріптердің (осы әріптен басталатын

сөздердің) мәтіндегі абсолюттік жиіліктері алдыңғы бағандағы реестрдегі кездесу санынан басқаша, олардың мәтіндегі кездесу жиіліктері есепке алынып, жиілік бойынша орындары ауысқан. Бұл статистикалық мәліметтер бойынша, қазақ тілінде мәтін ішінде сөз басында жиі қолданылатын әріп – **Б**, одан кейін – **Қ**. Одан кейін – **К, Ж, А, Т, С, Д, Е, О**. Осы 10 әріп мәтінде сөз басында жиі қолданыла келе 500 сөзқолданыстан тұратын мәтіннің 80 пайызға жуығын құрайды.

### 4.3.2 Қазақ тіліндегі сөздердің кері-әліпбилі жиілік сөздігі

Сөздіктердің басым көпшілігінде сөздер алдыңғы дыбыстар бойынша әліпби тәртібімен орналасса, кері-әліпби сөздікте, керісінше, сөз бен сөзтұлғалар соңғы әріп-дыбысынан басталып, әліпби тәртібіне келтіріледі және сөздердің соңғы жағынан тегістеліп беріледі. Сөздіктің бұл түрі «Кері-әліпби сөздік» немесе тек *«Кері сөздік»* деп аталып жүр. Егер белгілі бір мәтін бойынша түзілген мұндай сөздікте әр сөздің мәтіндегі қолдану жиіліктері де ескерілетін болса, ондай сөздікті «Кері-әліпби жиілік сөздік» деп атайды. Қазақ тілін зерттеу жұмыстарында мұндай сөздіктер де бірте-бірте қолданыс тауып келеді.

Кері сөздіктердің рөлі, әсіресе, агглютинативті тілдер үшін ерекше, алайда олардың саны әлі де болса санаулы. Осындай сөздіктердің алғашқылары ретінде: Қ.Бектаевтың «Қазақ тілінің кері алфавитті сөздігі» (Алматы, 1971), Р.Қунгуров, А.Тихонов «Обратный словарь узбекского языка» (Самарканд, 1969, өзбек тілінде) және т.б. кері сөздіктерді атауға болар еді. Аталған сөздіктер мәтінге тікелей қатыссыз, яғни бұрын жарық көрген түсіндірме, орфографиялық және т.б. сөздіктер негізінде жасалғандықтан, оларда сөздердің қолдану жиілігі көрсетілмеген.

Қазақ тілінің кері-әліпби жиілік сөздігінің ең алғаш жарық көруі М.О.Әуезов шығармаларының ЭЕМ арқылы зерттеле бастауымен тікелей байланысты. Бұл орайда 1979 жылы «Ғылым» баспасынан шыққан «М.Әуезовтің «Абай жолы» романының жиілік сөздігінің» және 1995 жылы «М.О.Әуезовтің 20 томдық шығармалар мәтіндерінің жиілік сөздіктерінің» жарық көргенін айрықша атап айтуға болады.

Жиілік сөздіктің кері-әліпби түрінде реестрлік сөзге (сөзтұлғаға) қосылған біркелкі жалғаулар мен жаңа сөз жасаушы немесе сөз түрлендіруші жұрнақтар бір жерге жинақталып берілуі тілді зерттеушілер үшін аса құнды тілдік мәліметтер болып табылады. Кері-әліпби жиілік сөздік морфологиялық категорияларды зерттеу кезінде де және сөз тудырушы, сөз түр-



Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
64	<i>астапыралла/од.</i>	2	3,504
65	<i>уа/од.</i>	2	3,511
66	<i>ә/од.</i>	4	3,525
-	-----	---	-----
100	<i>қуңқ/ел.</i>	1	4,326
101	<i>дүрк/ел.</i>	1	4,330
102	<i>сақ-сақ/ел.</i>	1	4,333
103	<i>таңқ/ел.</i>	1	4,336
104	<i>жарқ/ел.</i>	4	4,350
-	-----	---	-----
298	<i>кішкене/сн.</i>	29	12,121
299	<i>көне/сн.</i>	1	12,125
300	<i>сіре/сн.</i>	1	12,128
301	<i>кексе/сн.</i>	1	12,132
302	<i>келте/сн.</i>	1	12,135
-	-----	---	-----
1202	<i>қайда/ес.</i>	20	22,602
1203	<i>әлдеқайда/ес.</i>	2	22,609
1204	<i>ана/ес.</i>	13	22,654
-	-----	---	-----

**38-кестеде** жоғарыда сөз болған «Қазақ тіліндегі сөздердің кері-әліпбилі жиілік сөздігінен» үзінді берілді. Кері-әліпбилі сөздіктің анықтамасы бойынша және осы кестенің «Сөз/Сөз табы» деп аталатын 2-бағанасындағы сөздердің орналасуынан ең алғашқы байқайтынымыз, ол сөздердің соңғы әрпінен бастап, ең бірінші әрпіне қарай қатаң әліпби тәртібімен орналасуы. Сол себепті, сөз соңында тұрған сөз табын ажырату мақсатымен қойылған шартты белгіленімнің «а» әрпімен аяқталатын сан есімнің «-са» шартты белгісімен белгіленген сөздер кері-әліпбилі жиілік сөздіктің бас жағында тұратынын байқауға болады. Сөздік бойында мұндай орналасу барлық сөз таптары белгіленімдеріне де тән. Кері-әліпбилі жиілік сөздіктің анықтамасы мен құрылымы бойынша әр сөз табына қатысты сөздер өзара топ-топ болып орналасады. Осының нәтижесінде зерттеуші әр сөз табына қатысты сөздердің статистикасын жеке алып сипаттауына мүмкіндігі бар.

Қазақ тіліндегі сөздердің кері-әліпбилі жиілік сөздігіне статистикалық сипаттама беруге қатысы бар мәліметтер **38-кестедегі 1-бағанадағы** кері-

әліпбилі жиілік сөздіктің рет санына, **2-бағанадағы Сөз/Сөз табы** деп аталатын сөз табын ажырататын белгіленімдерге, **3-бағанадағы** сөздердің абсолютті жиілігіне және **4-бағанадағы** жеке сөз табына қатысты бір топ сөздердің қайталана қолдануларының нәтижесінде сол сөздің мәтін көлемін қамту пайызына қатысты мәліметтер берілген.

**39-кестеде** осы мәліметтерге статоллингвистикалық топтастырулар жасалады. Сөз таптарының реестр бойын және мәтін көлемін қамту пайыздары көрсетіледі.

**39-кесте. Кері-әліпбилі сөздігі бойынша жеке сөз табына қатысты сөздердің реестр бойын және мәтін көлемін қамту пайызы**

Рет саны ( <i>i</i> =)	Кері-әліпбилі жиілік сөздіктегі сөз табы және белгісі	Әр сөз табына қатысты сөздердің реестрдегі саны	Сөздердің реестр бойын қамту пайызы	Әр сөз табындағы сөздердің мәтіндегі қосынды абсолютті жиілігі	Әр сөз табына қатысты сөздердің мәтінді қамту пайызы
1	2	3	4	5	6
1.	<i>Сан есім –са.</i>	45	0,995	645	2,247
2.	<i>Модаль сөз –мд.</i>	16	0,354	357	1,243
3.	<i>Одағай сөз –од.</i>	34	0,752	83	0,289
4.	<i>Көмекші есім-ке.</i>	4	0,088	156	0,543
5.	<i>Еліктеуіш сөз –ел.</i>	51	1,128	94	0,327
6.	<i>Шылау –ил.</i>	88	1,946	1813	6,315
7.	<i>Сын есім –сн.</i>	963	21,301	3321	11,567
8.	<i>Есімдік –ес.</i>	86	1,902	2091	7,283
9.	<i>Үстеу –ұс.</i>	245	5,419	1134	3,950
10.	<i>Етістік –ет.</i>	1236	27,339	9947	34,646
11.	<i>Зат есім –зт.</i>	1753	38,775	9069	31,588
		$\Sigma=4521$	100%	$\Sigma=28710$	100%

**40-кесте. Кері-әліпбилі жиілік сөздік бойынша жеке сөз табына қатысты сөздердің реестр бойын қамту пайызы**

Рет саны (i=)	Кері-әліпбилі жиілік сөздіктегі сөз табы және белгісі	Әр сөз табына қатысты сөздердің реестрдегі саны	Сөздердің реестр бойын қамту пайызы	Бір топ сөз таптарына қатысты сөздердің реестр бойын қамту пайызы
1	2	3	4	5
1.	<i>Зат есім –зт.</i>	1753	38,775	38,775
2.	<i>Етістік –ет.</i>	1236	27,339	66,114
3.	<i>Сын есім –сн.</i>	963	21,301	87,415
4.	<i>Үстеу –ұс.</i>	245	5,419	92,834
5.	<i>Шылау –шл.</i>	88	1,946	94,780
6.	<i>Есімдік –ес.</i>	86	1,902	96,682
7.	<i>Еліктеуші сөз –ел.</i>	51	1,128	97,810
8.	<i>Сан есім –са.</i>	45	0,995	98,805
9.	<i>Одағай сөз –од.</i>	34	0,752	99,557
10.	<i>Модаль сөз –мд.</i>	16	0,354	99,911
11.	<i>Көмекші есім-ке.</i>	4	0,088	100%
<b>Қосындысы:</b> 4521		100%		

**41-кесте. Кері-әліпбилі жиілік сөздік бойынша жеке сөз табына қатысты сөздердің мәтін көлемін қамту пайызы**

Рет саны (i=)	Кері-әліпбилі жиілік сөздіктегі сөз таптары және олардың белгіленімі	Әр сөз табына қатысты сөздердің мәтін ішіндегі қосынды абсолютті жиілігі	Әр сөз табына қатысты сөздердің мәтінді қамту пайызы	Бір топ сөз таптарына қатысты сөздердің мәтін көлемін қамту пайызы
1	2	3	4	5
1.	<i>Етістік –ет.</i>	9947	34,646	34,646%
2.	<i>Зат есім –зт.</i>	9069	31,588	66,234%
3.	<i>Сын есім –сн.</i>	3321	11,567	77,801%

Рет саны (i=)	Кері-әліпбилі жиілік сөздіктегі сөз таптары және олардың белгіленімі	Әр сөз табына қатысты сөздердің мәтін ішіндегі қосынды абсолютті жиілігі	Әр сөз табына қатысты сөздердің мәтінді қамту пайызы	Бір топ сөз таптарына қатысты сөздердің мәтін көлемін қамту пайызы
4.	<i>Есімдік –ес.</i>	2091	7,283	85,084%
5.	<i>Шылау –шл.</i>	1813	6,315	91,399%
6.	<i>Үстеу –ұс.</i>	1134	3,950	95,349%
7.	<i>Сан есім –са.</i>	645	2,247	97,596%
8.	<i>Модаль сөз –мд.</i>	357	1,243	98,839%
9.	<i>Көмекші есім –ке.</i>	156	0,543	99,382%
10.	<i>Еліктеуіш сөз –ел.</i>	94	0,327	99,709%
11.	<i>Одағай сөз –од.</i>	83	0,289	100%
	<b>Қосындысы: 28710</b>	<b>100%</b>		

**39-кестедегі** жеке сөз табына қатысты сөздердің реестр бойын және мәтін көлемін қамту пайыздары жайлы мәліметтерді статистикалық сипаттау әрекеті ыңғайлы болу үшін оны екі кесте түрінде, яғни **40** және **41-кестелер** бойынша қарастыруды жөн көрдік.

**40-кестеде** жеке сөз табына қатысты сөздердің кері-әліпбилі жиілік сөздіктегі реестр бойындағы саны арқылы олардың реестрдегі үлесі пайыздық салмақ арқылы анықталды. **40-кестеде** көрініс тапқандай, әр сөз табына қатысты сөздердің саны әртүрлі болуынан біз олардың аз, орташа немесе көп кездесуінің статистикалық сипатын айта аламыз. Мысалы, біз қарастырып отырған кері-әліпбилі жиілік сөздіктің реестрлік тізімінде 4521 сөз болса, олардағы 11 сөз табына қатысты сөздердің тек реестрден алатын пайыздық салмағы ғана емес, сонымен бірге мәтін ішінде қайталана қолдана келіп, мәтінді қамту пайызының да аз не көптігін сөз етуге болады. Сол сияқты, тек әр сөз табының қолданылуы ғана емес, олардың топтарының да статистикасын сипаттай аламыз.

Мысалы, **40-кесте** бойынша **зат есімге** қатысты сөздердің реестрде **2 мыңға** жуық кездесіп (**1753 рет**), реестр бойының **38,775 пайызын** қамтып жатқанын (немесе шамамен **39 пайызын**) байқаймыз. Сол сияқты, **етістік сөз табына** қатысты реестрлік сөздерді де жеке айтатын болсақ, олар да реестр бойында зат есімнен кейін ең көп кездесетін реестрлік сөздер қатарына жатады, яғни олардың саны **1236 тең** және реестр бойының **27,339 пайызын**



қамтиды екен. Бір сөзбен айтқанда, аталған зат есім мен етістік сөздерді біріктіріп қарастырсақ, олар кері-әліпбилі жиілік сөздіктің реестрінде ең көп кездесіп, барлық реестр бойының **66-дан** аса пайызын қамтып жатқанына көз жеткізуге болады. Осылайша **40-кесте** мәліметтерін үш топқа бөліп қарастырсақ, мысалы, реестрде сан жағынан ең көп кездесетін **1-топқа**: *Зат есім (зт.), Етістік (ет.), Сын есім (сн.), Үстеу (үс.);* реестрде сан жағынан орташа кездесетін **2-топқа**: *Шылау (шл.), Есімдік (ес.), Еліктеуіш сөз (ел.), Сан есім (са.);* реестрде сан жағынан ең аз кездесетін **3-топқа**: *Одағай сөз (од.), Модаль сөз (мд.), Көмекиі есім (ке.).*

Енді осы аталған топтардың реестр бойын қамту пайызына тоқталайық. **1-топтағы**: *зат есім, етістік, сын есім, үстеу* сөздер реестр бойының **92,834 %**, **2-топтағы**: *шылау, есімдік, еліктеуіш сөз, сан есім* сөздер реестр бойының **5,971 %**, **3-топтағы**: *одағай сөз, модаль сөз, көмекиі есім* ең аз кездесетін сөздер реестр бойының **1,194 пайызын** қамтып жатыр екен.

**41-кестедегі** реестрдегі әр сөз табына қатысты сөздердің қайталана қолдануының қосынды абсолютті жиіліктерінің мәтін көлемін қамту пайызы жайлы мәліметтерге статистикалық сипаттама берейік.

**41-кестедегі** статистикалық деректердің негізінде сипаттама жазу, яғни әр сөз табына қатысты сөздердің қайталана қолданып, реестр бойын емес, мәтін көлемін қамтудың статистикасын шығару үшін ең алдымен саны 11-ге тең сөз таптарын олардың жиі, орташа, сирек қолдануларына қарай үш топқа бөліп қарастыруды жөн санадық. Сондықтан **41-кестедегі** статистикалық деректерге сүйене отырып, **1-топқа** ең жиі қолданған 9947 рет қолданып, мәтіннің **34,646** пайызын қамтып, жиілік сөздіктің 1-ші орнына ие болып отырған **етістік сөз табын** және абсолютті жиілігі **9069-ға** тең, ал мәтінді қамту пайызы **31,588-ге** тең жиі қолдану жағынан екінші орынды алып тұрған **зат есім** сөз табын **1-топқа** жатқызуды ұйғардық. Ең жиі қолданыстағы екі сөз табынан-ақ тұратын бұл **1-ші топ** мәтіннің **66,234%** қамтиды. Сол сияқты, **5 сөз табына** қатысы бар орташа жиілікті сөздерден тұратын, яғни **сын есім, есімдік, шылау, үстеу, сан есімге** қатысты сөздер **2-топты** құрайды және олар қайталана қолданып, мәтіннің **31,362**. Енді ең сирек жиілікті қолданыстағы соңғы **3-топқа** келесі сөз таптары жатады: **модаль сөз, көмекиі есім, еліктеуіш сөз, одағай сөздер**. Аталған **3-топтағы** сөздер қайталана қолдана келе, барлық мәтіннің тек болғаны **2,402 пайызын** ғана қамтитыны дәлелденді.

Сонымен, сөздердің кері-әліпбилі жиілік сөздігі бойынша сөз таптарының реестр бойын қамтуына қарай зат есім сөздер көп кездеседі. 4521 реестр сөзден тұратын бұл жиілік сөздікте зат есімдер 1753 реестр сөзді, ал

егістіктер 1236 сөзді құрайды. Екеуі қосылып, ресстр бойынның 40 пайызға жуығын алады екен. Одан кейінгі орында сын есімдер тұр.

Ал сөз таптарының мәтінде қайталана қолданыла келе, абсолютті жиіліктері бойынша жоғарыдағы сөз таптарының жиілік дәрежесі өзгереді. Мәтінде жиі қолданылуы бойынша егістіктер зат есімдерді басып озып, алдыңғы орынға шығып кетеді. Демек, ресстр бойын қамту пайызы жағынан зат есімдер алдыңғы орында, мәтінді қамту пайызы бойынша, егістіктер алдыңғы орында, ал екеуінде де сын есімдер үшінші орынды иеленеді. Модаль сөздер, көмекші есімдер, еліктеуіш сөздер, одағай сөздер сирек кездесетіндерге жатады.

### 4.3.3 Қазақ тіліндегі сөздердің жиілікті әліпбилі жиілік сөздігіне қатысты статистикалық сипаттама

Тілдік фактілерді сандық ұғымдарға жүгініп баяндау – қазақ тіл білімін зерттеуші ғалымдарына да тән. Өйткені қай кездегі болмасын қазақ тілі жайлы зерттеулердің бәрінде дерлік *аз, көп, көптеген, жиі, сирек, өнімді, өнімсіз, мол, жуық, тең, шамалас, көп рет* т.б. шама, мөлшерді көрсететін сөздер кездесіп отырады және бұлардың мәнін нақты санмен көрсетуге талаптанушы ғалымдар да бар. Сөйтіп, қолданбалы тіл білімінің бүгінде «Статистикалық лингвистика» деп аталып жүрген жаңа саласының алғашқы нышандары мен белгілері, міне, осылардан басталады.

Сөздіктердің «жиілік сөздік» түрі Әліпбилі жиілік сөздіктен басқаша, дәлірек айтқанда, әр сөздің (сөзтұлғаның) қолдану жиілігінің дәрежесіне қарай орналасады: ең бірінші ретте орналасатын мәтіндегі ең жиі қолданылған сөз (не сөзтұлға), екінші, үшінші (тағы тағылар) ретте орналасатын сөздер – кездесу жиіліктері бірте-бірте кеміп отыратын сөздер (сөзтұлғалар). Егер бірнеше сөздің (не сөзтұлғаның) жиіліктері тұрақтылық сипатта болса, яғни абсолютті жиіліктері бір-біріне тең болып келсе, онда ол сөздер (сөзтұлғалар) жиілік сөздікте әліпби тәртібімен реттеледі.

Ана тіліміздің немесе басқа да туыстас (не туыстас емес) тілдердің әр стильдік мәтіндеріне қатысты жасалған **жиілік сөздіктер** сол тілдің лексикалық құрылымын терең зерттеуде, лексиканың жалпы функционалдық стильге ортақ бөлігін ажыратуда септігі мол.

Сондай-ақ, жалпыхалықтық лексиканың, кірме сөздер мен терминдердің, неологизмдердің және басқа лексикалық топтардың тілдің жалпы лексика жүйесінен алатын орны жөнінде де жиілік сөздік бойынша көптеген сандық және сапалық мәліметтер алуға болады. Сөздердің саны мен тұлғалары

арқылы тілдің лексикалық байлығы мен стильдік ерекшеліктері анықталады. Мұндай сөздіктерден зерттеуші жаңа сөзжасам жүйесіне қатысты да бағалы деректер ала алады. Сонымен бірге, жиілік сөздік мәліметтері қазақ тілін туыстас түркі тілдері және туыс емес тілдермен салыстыру – типологиялық зерттеу үшін де аса қажетті екенін атап өткен жөн.

Жиілік сөздіктер күнделікті қолданбалы сипаттағы міндеттерді шешуде де маңызды. Мәселен, әртүрлі типтегі минимум сөздіктерін жасауда және шет тілдерін (екінші тілді) үйрену мен оқытуға қажетті статистикалық әдістерді қолдануда да қажетті. Сонымен бірге жиілік сөздіктердің тілдік және сандық мәліметтері оқулықтар мен оқу құралдарын құрастыру ісінде тілдік материалдарды тиімді түрде орналастыруда және әркілы лексикалық топтар мен грамматикалық категорияларды орынды түрде оқыту процесінде де таптырмас тілдік құрал рөлін атқара алады.

Жиілік сөздіктер материалдарының тіл білімін зерттеудегі қажеттілігі мен бағалылығы жөнінде Л.Н.Засорина мынадай пікір айтқан болатын: «Материалы частотных словарей исключительно ценны и для собственно лингвистических исследований. Они оказывают влияние на судьбы традиционной лексикографии. Ими пользуются в решении основной проблемы общей лексикологии – выделение словарного фонда активного и периферического словаря; они полезны и при изучении вопросов стилистики, семантики и литературной нормы языка».

Қазақ тіліндегі сөздердің жиілікті әліпбилі жиілік сөздігіне қатысты статистикалық сипаттама берер алдында ондай сөздіктің *42 -кестедегі* үзіндісімен таныс болайық.

**42-кесте. Қазақ тіліндегі сөздің жиілікті әліпбилі сөздігінен үзінді**

Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
1	2	3	4
1	<i>бол/ет.</i>	666	2,320
2	<i>де/ет.</i>	541	4,204
3	<i>еді/ет.</i>	479	5,873
4	<i>кел/ет.</i>	407	7,290
5	<i>бір/са.</i>	352	8,516
6	<i>ал/ет.</i>	335	9,683
7	<i>да/шл.</i>	334	10,846
8	<i>бұл/ес.</i>	299	11,888



Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
43	<i>қыл/ет.</i>	99	31,240
44	<i>бар/мд.</i>	98	31,581
45	<i>жақ/зт.</i>	97	31,919
46	<i>ат/зт.</i>	94	32,247
47	<i>екі/са.</i>	91	32,564
48	<i>ет/ет.</i>	90	32,877
49	<i>сал/ет.</i>	89	33,187
50	<i>қой/ет.</i>	88	33,494
51	<i>не/ес.</i>	86	33,793
52	<i>алды/ке.</i>	84	34,086
53	<i>ел/зт.</i>	84	34,378
54	<i>баста/ет.</i>	81	34,660
55	<i>өт/ет.</i>	81	34,943
56	<i>кісі/зт.</i>	79	35,218
57	<i>бірақ/шл.</i>	75	35,479
58	<i>мен/ес.</i>	74	35,737
59	<i>кір/ет.</i>	73	35,991
60	<i>ақ/сн.</i>	72	36,242
61	<i>қол/зт.</i>	71	36,489
62	<i>енді/үс.</i>	67	36,722
63	<i>жол/зт.</i>	67	36,956
64	<i>адам/зт.</i>	64	37,179
65	<i>бет/зт.</i>	63	37,398
66	<i>күл/ет.</i>	63	37,618
67	<i>қара/сн.</i>	63	37,837
68	<i>жалғыз/сн.</i>	61	38,049
69	<i>жан/зт.</i>	61	38,262
70	<i>мал/зт.</i>	61	38,474
71	<i>ғой/шл.</i>	60	38,683
72	<i>қас/зт.</i>	60	38,892
73	<i>тап/ет.</i>	60	39,101
74	<i>аз/сн.</i>	59	39,307
75	<i>үсті/ке.</i>	59	39,512
76	<i>жігіт/зт.</i>	58	39,714



Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
111	<i>көш/зт.</i>	38	45,423
112	<i>тау/зт.</i>	38	45,556
113	<i>әке/зт.</i>	37	45,684
114	<i>жай/зт.</i>	37	45,813
115	<i>қазак/зт.</i>	37	45,942
116	<i>қала/зт.</i>	37	46,071
117	<i>қар/зт.</i>	37	46,200
118	<i>тай/зт.</i>	37	46,329
119	<i>ара/зт.</i>	36	46,454
120	<i>аш/ет.</i>	36	46,580
121	<i>барлық/ес.</i>	36	46,705
122	<i>қызыл/сн.</i>	36	46,830
123	<i>сен/ес.</i>	36	46,956
124	<i>су/зт.</i>	36	47,081
125	<i>жақсы/сн.</i>	35	47,203
126	<i>орта/зт.</i>	35	47,325
127	<i>соқ/ет.</i>	35	47,447
128	<i>есті/ет.</i>	34	47,565
129	<i>жұрт/зт.</i>	34	47,684
130	<i>қайт/ет.</i>	34	47,802
131	<i>өзге/сн.</i>	34	47,921
132	<i>таста/ет.</i>	34	48,039
133	<i>күй/зт.</i>	33	48,154
134	<i>мін/ет.</i>	33	48,269
135	<i>ой/зт.</i>	33	48,384
136	<i>сияқты/шл.</i>	33	48,499
137	<i>сұра/ет.</i>	33	48,614
138	<i>ауыз/зт.</i>	32	48,725
139	<i>кітап/зт.</i>	32	48,837
140	<i>құлынишақ/зт.</i>	32	48,948
141	<i>ма/шл.</i>	32	49,060
142	<i>өл/ет.</i>	32	49,171
143	<i>сөйле/ет.</i>	32	49,282
144	<i>аяқ/зт.</i>	31	49,390





Рет саны №	Сөз/Сөз табы	Абсолютті жиілік	Бір топ сөздің мәтінді қамту пайызы
4510	<i>іркiл/ет.</i>	1	99,962
4511	<i>ірікте/ет.</i>	1	99,965
4512	<i>ірілік/зт.</i>	1	99,969
4513	<i>іріткі/зт.</i>	1	99,972
4514	<i>іс/ет.</i>	1	99,976
4515	<i>істік/зт.</i>	1	99,979
4516	<i>іші-бауыр/зт.</i>	1	99,983
4517	<i>ішіп-же/ет.</i>	1	99,986
4518	<i>экономикалық/сн.</i>	1	99,990
4519	<i>январь/зт.</i>	1	99,993
4520	<i>япырай/од.</i>	1	99,997
4521	<i>япырау/од.</i>	1	100
	<b>Барлығы:</b>	28710	

Сөздік пен мәтін арақатынасын анықтаудың ең бір қолданбалылық маңызы зор түрі – жиілік сөздіктегі сөздің не сөздер тобының мәтін бойын қамту дәрежесін айқындау. Мәселен, шет тілін не ана тілін үйренуде не үйретуде минимум сөздіктер жасау қажеттігі туатыны белгілі. Минимум сөздікке енетін сөздер сан жағынан аз болғанымен, мәтін ішінде көп қайталанып барып, оның өн бойын қамту мүмкіндігі мол. Немесе тілімізде кейбір сөздер сирек қолданылғанымен, ондай сөздерді ескермеу, мәтін мазмұнының (сөйлем мағынасының) дұрыс ашылмауына әкеліп соғуы ықтимал.

Жиілік сөздіктің құрылымы бойынша жиі қолданыстағы сөздер сөздік бойының бас жағында орналасатындығын ескеріп, олардың мәтін мен сөздікті қамту пайызы жайлы мәліметтер **43-кестеде** берілді.

**43-кесте.** Бір топ сөздің мәтін мен сөздік бойын қамту пайызы

Топтар рет саны	Рангтық шекаралық	Сөздік бойын қамту пайызы	Мәтін бойын қамту пайызы
1	2	3	4
1	1	0,022	2
2	1– 6	0,132	10
3	1 – 19	0,420	20
4	1 – 40	0,885	30



Енді **44-кестедегі** жиілікті әліпбилі сөздік бойынша бірдей жиілікті сөздердің реестр бойы мен мәтін бойын қамту пайызына қатысты статистикалық мәліметтерге сипаттама берейік.

**44-кесте. Жиілікті әліпбилі сөздік бойынша бірдей жиілікті сөздердің реестр мен мәтін бойын қамту пайызы**

Сөз жиілігі	Реестрдегі сөз саны	Сөздік бойын қамту пайызы (%)	Жиынтық абсолютті жиілік	Мәтін бойын қамту пайызы (%)
1	2247	49,701 %	2247	16,517 %
2	709	15,682 %	1418	4,939 %
3	359	7,941 %	1077	3,751 %
4	231	5,109 %	924	3,218 %
5	168	3,716 %	840	2,926 %
6	106	2,345 %	636	2,215 %
7	79	1,747 %	553	1,926 %
8	79	1,747 %	632	2,201 %
9	55	1,216 %	495	1,724 %
10	45	0,995 %	450	1,567 %

Бірдей жиілікті сирек қолданатын сөздерді (сөзтұлғаларды) жеке қарастырсақ, ең бірінші байқалатын жайт – жиілігі кеміген сайын ондай сөздердің санының да өсуі. Мәселен, ең сирек қолданыстағы сөздердің кездесу жиілігінің кемуі (азаюы) олардың сөздіктің бірлігі ретіндегі сандық мөлшерінің ұлғаюына да әкеліп соғады. Сондықтан **1 рет** қолданыста болатын сөздердің саны **2 реттік** қолданыстан әрі көп, әрі мәтін мен сөздік бойын қамту жағынан да алдында тұрады. Сол сияқты, жиіліктері **2 реттен** болатын сөздер саны мен мәтінді қамту мүмкіндігі **3 реттегіге** қарағанда көп. Мысалы, жиілігі **1-ге тең** сөздер біз қарастырып отырған жиілік сөздіктің реестрлік қатарында 2247 рет орын алған, ал бұл сан бірдей жиілікпен кездесетін басқа сөздерге қарағанда ең көбі деуге болады және бұл сан – 2247-ге тең және реестрдегі барлық 4521 сөздің 49,701 пайызын, яғни барлық реестрлік сөздердің жартысын қамтып жатыр деуге болады. Сонымен бірге аталған 2247 сөз бір реттен ғана қолданыла отырып, мәтін көлемінің 16,517 пайызын қамтитынын жиілік сөздіктен анықтауға болады.

Сол сияқты, жиіліктері 2-ге, 3-ке, 4-ке, ... 10-ға және т.б. бұлардан 1 санына өсіп отыратын жиіліктер санына тең болып қолданылған реестрлік сөздердің кездесу саны өскен сайын, олардың реестр бойын және мәтін

көлемін қамту пайыздары міндетті түрде азая түсетіні тек қазақ тілінің заңдылығы ғана емес, барлық тілдерге ортақ заңдылық деуге болады.

Сонымен, сөздердің жиілікті әліпбилі сөздігі бойынша, **42-кестеде** мәтіннің 50 пайызын қамтитын жиі қолданылатын сөздер тізімі (150-ші топ деген сөзге дейін) толықтай берілді, ал 60, 70, 80, 90, 100 пайызын қамтитын сөздер үзік түрінде қысқартылған. **43-кестеде** жиілігінің кему тәртібіне қарай сөздер 10 топқа топтастырылып берілген. Ең алғашқы топқа жиілік сөздіктегі ең жиі кездесетін «*бол*» етістігі мәтін ішінде **666 рет** қолданып, жалғыз өзі-ақ мәтіннің **2,320** пайызын қамтыған. *Бол* етістігінің ең жоғарғы жиілікті көрсетуі бұдан бұрын шыққан сөздіктерде де орын алған. **Бол** етістігінің жоғары жиілікті көрсетуі оның көмекші етістік болып, грамматикалық мағына үстеуші қызметімен тікелей байланысты деуге болады. Өйткені жиі қолданылу, грамматикалық абстракция жасайтын грамматикалық бірліктерге тән қасиет. Жиілік сөздікте *бол* етістігі сияқты жоғары жиілікте тұрған көп сөз осы көмекші етістіктер. Мысалы, *де, еді, кел, ал, тұр, отыр, жатыр, көр, айт, бер, шық, кет, жүр, бар, түс, қыл, сал, қой, баста* т.б. Сол сияқты жиілік сөздікте шылаулардан *да, де, мен, бірақ, зой, қарай, ғана, және, тағы* шылаулары, сондай-ақ есімдіктер (*бұл, өз, сол, ол, осы, бар, мен*) жиі кездескен. Бұл жиілік сөздік тек көркем шығарма мәтіндерінен алынғандықтан, термин сөздер аз, тіптен кездеспейді десе де болады.

## ҚОРЫТЫНДЫ

Әріптер мен әріп тіркестерінің жиіліктерін анықтау қазақ әліпбиін латын қарпіне көшірудегі бірнеше теориялық-практикалық бағыттағы мәселелерді шешуге септігін тигізеді. Әсіресе, әріп, әріп тіркестерінің жиілігі мәтінді қолмен теруде жылдамдықты арттыру үшін латын қаріпті қазақ пернетақта-сына әріптерді тиімді орналастыруда қажет.

Қазақ тіл білімінде әріптерді статистикалық тұрғыдан зерттеудің тарихы 1973 жылы жарық көрген «Қазақ тексінің статистикасы» атты еңбектен басталады деуге болады.

Кітапта қазақ тіліндегі әріп жиілігі бойынша жасалған зерттеулерге шолу жасалып, қазақ тіліндегі жиі кездесетін дыбыстар (әріптер) мен сирек кездесетін әріптердің жиілігі бойынша сипатталды. Зерттеулердегі жанрлық, екі тілдегі (қазақ және өзбек говорлары) әріптердің жиілігі бойынша салыстырмалы талдаулардан әріп жиілігі бойынша негізінен ұқсас, барабар жиынтық шамалар алынғанын байқаймыз.

Ұсынылып отырған қазақ әріптерінің жиілік сөздігі А.Байтұрсынұлы атындағы Тіл білімі институтының «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» мегажобасы бойынша орындалған «Жалпы білім берудегі қазақ тілінің жиілік сөздігінің» бір түрі ретінде алынған үлкен бір нәтиже – «Сөзформалардың жиілік сөздігінің» мәтінінен алынды. Қазақ әріптерінің жиілік сөздігі екі түрлі нұсқада жасалды. Бірі – «Қазақ тіліндегі әріптердің Әліпбилі жиілік сөздігі», екіншісі «Қазақ тіліндегі әріптердің жиілікті әліпбилі жиілік сөздігі».

Бұның алдында жасалған әріптердің статистикалық талдауларындағыдай, қазақ тілінде ең жиі кездесетін әріп – **А** фонемасы. Одан кейінгі орында **Е**, **Ы** фонемалары. Ал **Ө**, **Ұ**, **Ү** дыбыстарының сирек кездесуі қазақ тілінде ерін үндестігінің жазуда сақталмауынан болар деп ойлаймыз. Яғни бұл дыбыстар түбір сөздер құрамында болғанмен, қосымшалар құрамында езулік дыбыстармен жазылғандықтан, аз санды көрсетіп отыр. Ал дауыссыздар ішінде ең жиі кездесетін әріп – **Н** фонемасы, **Т**, **Р**, **Л** әріптері жиілігі бойынша екінші орында. 30-орында тұрған **Ә**-ден төмен тұрған әріптердің барлығы (я, х, ц, ф, в, э, ь, ю, һ, ь, ч, щ, ё) кірме әріптер. **Ә** әрпінің сирек кездесуі бір жағынан қосымшаларда кездеспеуінен болуы да мүмкін.

Қазақ тіл білімінде әріп тіркестерін статистикалық тұрғыдан зерттеудің тарихы да 1973 жылы жарық көрген «Қазақ тексінің статистикасы» атты еңбектен басталады. Бұл ретте Қ.Бектаев, А.Қ.Жұбановтың «Графема-фонема тіркестерінің жиілік тізімдері туралы», А.Қ.Жұбановтың «Қазақ текстеріндегі әріптер тіркестерінің статистикасы», Е.Агманов пен А.Қ.Жұбановтың «Распределения частот появления сочетаний знаков в орхоно-енисейской письменности», Қ.Бектаевтың «Қарақалпақ тіліндегі фонема тіркестерінің статистикасы» атты еңбектерін атап айтуға болады.

Ұсынылып отырған қазақ әріптер тіркесінің жиілік сөздігі жоғарыда аталған «Сөзформалардың жиілік сөздігінің» мәтінінен алынды.

Қазақ әріп тіркестерінің жиілік сөздігі екі түрлі нұсқада жасалды. Бірі – «Қазақ тіліндегі әріп тіркестерінің Әліпбилі жиілік сөздігі», екіншісі – «Қазақ тіліндегі әріп тіркестерінің жиілікті әліпбилі жиілік сөздігі».

Екіәріптік жиілік сөздікте әртүрлі тіркестердің рет саны 1160-қа тең, ал осы екіәріптік тіркестер қайталанып қолданыла келіп, олардың абсолютті жиіліктерінің қосындысы 67 604 670 мәтін бойындағы екіәріптік тіркестердің 100 пайызын құрайды. Ең жиі тіркесетін әріп тіркестері ретінде *ар, ан, ал, ын, да, та, ен, ер, қа, де, ін, ла, ды, ты...* атауға болады. Осы сөздіктің жоғары жиілікті бөлігінен үзінді ретінде берілген 217 екіәріптік тіркес барлық мәтіннің 90 пайызын қамтиды. Статистикалық мәліметтер толық болу үшін біз өз зерттеуімізде сөз басында және сөз соңында кездесетін екіәріптік тіркестерінің де жиілік сөздіктерін құрастырдық.

Әліпби ауыстыруда маңызды мәселелердің бірі – әріптерді пернетақтада орналастыру. Әріптерді пернетақтада орналастыруда қандай ұстанымдарға сүйенеміз деген мәселенің басын ашып алу қажет. Оның *біріншісі* – пернетақтадағы ағылшынша жазатын латын қаріптерінің тұрған орнын сақтап, кирилдегі қазақша әріптер тұрған жоғарғы жақ қатарға жаңа таңбаларды орналастыру. Бұл ағылшын пернетақтасында тұрған латын қаріптерін басуда шатастырмау үшін тиімді деуге болады. *Екіншісі* – ағылшын пернетақтасындағы латын қаріптерін сақтай отырып, жоғары жаққа емес, негізгі үш қатардағы бос орындарға жаңа таңбаларды қою. *Үшіншісі* – әлемдік тіл білімінде ұстанылатын принцип – пернетақтада әріптерді қолданыста жиі кездесуіне қарай орналастыру, яғни жиі кездесетін әріптерді пернетақтада басқанда оңай, ікемді басылатын жерлерге қою болып табылады.

Әріптерді пернетақтада тиімді орналастыру мәселесіне қатысты А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімі қолданыстағы әріптердің кездесу жиілігін анықтады. Осы жиілік талдау (әріп және әріп тіркесінің жиілік сөздіктері) нәтижесі бойынша қазақ әріптері мен әріп тіркестерінің жиі кездесуі ескеріліп орналастырылған пернетақтаның пилоттық жобасы ұсынылды.

Пернетақтада әрбір әріптің кирилдегі және латындағы баламалары келтірілді. Пернетақтаның ортаңғы бөлігінде жиілік сөздіктегі ең жиі қолданылған әріптер орын алды, ал олардың оң жақтарында екіәріптік тіркес жиілік сөздігінен алынған әріптер орналастырылды. Пернетақтаның шеткі жақтарына сирек кездескен әріптер орналастырылды. Осы әріптер орналастырылған пернетақтада теруге арналған арнайы компьютерлік бағдарлама жасалды. Жиілік талдау бойынша орналастырылған латын әріптері бірнеше пернетақта батырмасына жапсырылып, бұл пернетақта нұсқасының мәтін теруге ыңғайлы, тиімділігі уақыт пен сапасына қарай сынақтан өткізілуде.

Фонетика саласындағы буын теориясы ең күрделі және маңызды мәселелердің бірі деуге болады. Буынның жиілік сөздігін алу басқа тілдік бөліктерге қарағанда күрделірек. Өйткені қазақ тілінде сөздерді буынға бөлу айтылымда жақсы ажыратылғанмен, кирилл қарпінің әсерінен көптеген сөздерімізді буын жігін сақтамай жазу нормаға түскендіктен, ауызша буын-

ды басқаша дыбыстаймыз, жазуда басқаша жазамыз. Сондықтан сөздерді буынға бөлу жазуға бағына бермейді.

Қ.Жұбановтың «Буын жігін қалай табуға болады?» атты мақаласында буындардың моделі арқылы «буын жігін оңай, механик түрде табуға болады» – дейді. Қазақ тіл білімінде белгілі ғалым Қ.Бектаев қазақ буындарының жиілік сөздігін 1973 жылы шыққан «Қазақ тексінің статистикасы» атты жинақта береді.

Қазақ буындарының жиілік сөздігін қазіргі заманауи және қазақ тілінің барлық стильдерін қамтитын мәтіндерден алу мақсатында «Қазақ тіліндегі буындардың жиілік сөздігін» ұсынып отырмыз. Бұл жиілік сөздікті жасауда буын жігін автоматты жолмен ажырату қиын болғандықтан, Қолданбалы лингвистика бөлімінің серверде орналастырылған «Қазақ тілінің мәтіндер корпусынан» қазақ тілінің бес стилі бойынша жасалған 6 миллион сөзқолданыстан тұратын мәтіндер жинағынан әр стиль бойынша 10 000 сөзқолданыстан, барлығы – 50 000 сөзқолданыстан тұратын мәтін алынып, «қолдап» буынға бөлінді, яғни «Қазақ тілі буындарының жиілік сөздігін» құрастырудан бұрын көлемі 50 мың сөзқолданыстан тұратын мәтін бойынша сөзформалардың жиілік сөздігі құрастырылды. Бұл сөзформалардағы сөзтізбенің рет саны 32304 сөзформаға тең.

Бұл «Қазақ тіліндегі буындардың жиілік сөздігі» дәстүрлі типі бойынша үш түрлі нұсқада жасалды. «Қазақ тіліндегі буындардың Әліпбилі жиілік сөздігі», «Қазақ тіліндегі буындардың жиілікті әліпбилі сөздігі» «Қазақ тіліндегі буындардың кері-Әліпбилі жиілік сөздігі».

Кітапта берілген «Қазақ тілі буындарының жиілік сөздігінде» қазақ тіліндегі буындардың реттік саны – 1225, олардың қайталана қолданылуы негізінде буындардың қосынды жиілігі 124063-ке тең. Буынның жиілік сөздігі бойынша, қазақ тілінде сан жағынан бітеу буындар (үш және төрт әріпті) көп екен. Жоғарыда 1225 реестрдің 952-сін құрап, 40 пайызын осы бітеу буындар алады. Бітеу буындардың екі моделі бойынша пайыздық көрсеткішін анықтайтын болсақ,  $C+V+C+C$  төрт әріпті бітеу буын – 0,24 %,  $C+V+C$  үш әріпті бітеу буын – 40,753%. Алайда бітеу буындардың саны көп болғанмен, ашық буындар кездесу жиілігі жағынан бітеу буындарды басып озады. Мәселен, жоғарыдағы буындардың жиілікті әліпбилі сөздігіндегі жоғары жиіліктегі 26 буынның 25-і ашық буындар. Ашық буындар реестр саны бойынша 173 болғанмен, мәтінде қайталана қолданылып, мәтіннің 55 пайызға жуығын құрайды. Ал тұйық буындар реестр саны бойынша 100-ге жетпейді және жоғары жиіліктегі 26 буын реестрінің құрамында кездеспейді. Мәтінде қайталанып қолданыла келе, мәтіннің 4,24% пайызын ғана құрайды. Үш әріпті тұйық буындарға қарағанда екі әріпті тұйық буындар көп, жиілік сөздікте үш әріпті тұйық буындар 9 ғана болса, 91-і екі әріпті тұйық буындар, екеуі қосылып, 100 реестрлік қатарды құрайды.

Қайсібір тілде болмасын сөздер жалғыз өздері тұрып, бір-бірімен байланыспаса, ой дұрыс жеткізілмейді. Сөз бен сөзді байланыстырудың құралы тілдегі грамматикалық бірліктер болып табылады. Грамматикалық бірліктер (қосымшалар, көмекші сөздер) түбір сөздерге жалғанып не тіркесіп келіп, сөздерді байланыстырады. Тілдегі лексемалар сияқты қосымшалардың да

байланысу тәсілдері мен қолданыс аясы бар. Сөз түрлендіруші қосымшалар әдетте сөздерге талғамай жалғанады да, сөздерді түрлендіреді. Осындай сөз түрлендіруші қосымшалардың саны (варианттарын қосқанда) мен түрленім формалардың санын және олардың мәтіндердегі қолданыс жиілігін анықтаудың маңызы зор.

Ә.Ахабаев пен Қ.Бектаев зат есімдердің, Ә.Ахабаев есімдіктердің морфологиялық құрылымына талдау жасаса, С.Мырзабеков туынды түбір туынды етістіктердің, Қ.Бектаев пен С.Мырзабеков етістіктің форма өзгерткіш аффикстеріне, А.Белботаев ғылыми-техникалық стильдегі сын есім аффикстеріне, А.Жұбанов пен Е.Жұбанов «Абай жолы» романындағы сын есімдердің лексика-морфологиялық формаларына статистикалық талдаулар жасады.

Қазақ тіліндегі түрленім қосымшаларының, яғни форма тудырушы қосымшалардың жиілігі Қолданбалы лингвистика бөлімі құрастырған «Қазақ тілінің мәтіндер корпусы» деп аталатын 6 миллионнан тұратын бес стильден алынған корпус мәтіндерінен алынды.

Мәтіндерден қосымшалар жиілігін алу үшін мәтін алдымен морфологиялық анализатордан өткізілді. Морфологиялық анализатор автоматты түрде мәтіндегі сөздерді түбір мен қосымшаға бөліп, содан соң лемматизацияланған (түбірі бөлінген) бөліктен бөлек екінші жағындағы сөзтүрленім формаларын арнайы қосымшалар кестесі бойынша бөлшектеп және кестеде берілген шартты грамматикалық белгіленімдері бойынша сипаттап шығады. Содан кейін бірдей қосымша бөліктерін грамматикалық сипаттамаларымен бірге топтастырып, бірдей формалардың санын шығарады. Ұсынылып отырған қосымшалар сөздігі осылай автоматты әдіспен алынып отыр. Автоматты әдіс болғандықтан, омонимдерге қатысты кейбір нақты емес жағдайлар кездесуі мүмкін. Алайда 95 пайыз дұрыс ақпарат береді деп ойлаймыз.

«Қазақ тіліндегі қосымшалардың жиілік сөздігі» әдеттегідей үш типте жасалды: «Қазақ тіліндегі қосымшалардың Әліпбилі жиілік сөздігі», Қазақ тіліндегі қосымшалардың жиілікті әліпбилі сөздігі», «Қазақ тіліндегі қосымшалардың кері-Әліпбилі жиілік сөздігі».

Біз қарастырған мәтіндерде түрленім қосымшалардың 26 түрі айқындалды. Жоғарыдағы статистикалық мәліметтерді қорытындыласақ: қосымша атауларының 12-сі жоғары жиілікте кездеседі. Олар: *тәуелдік жалғау, табыс септік, есімше, көсемше, жіктік жалғау, барыс септік, көптік жалғау, жатыс септік, ілік септік, тұйық етістік, шығыс септік, болымсыздық жұрнақ* түрлері қайталана қолданыла келіп, мәтіннің 93,74 пайызын қамтыған. Осы қосымша атаулары ішінде жиілігі жағынан *тәуелдік жалғау, табыс септік, есімше, көсемше* түрлері мәтіннің 55,33 пайызын қамтып, қазақ тілінде ең жиі кездесін қосымшалар екендігін көрсетеді.

*Түрленім формалар дегеніміз* – сөздің түбіріне (негізгі және туынды) жалғанатын қосымшалардың жеке немесе бірінің үстіне бірі жалғануындағы сөздің түрленуі.

Түрленім формалардың жиілік сөздігі де жоғарыда аталған 6 миллиондық корпус мәтіннен автоматты жолмен алынды. Олар: «Қазақ тіліндегі түрленім формаларының Әліпбилі жиілік сөздігі», Қазақ тіліндегі түрленім





70-жылдары Қ.Бектаевтың басшылығымен және жетекшілігімен А.Қ.Жұбанов, С.Мырзабеков, А. Белботаев, Ә.Ахабаев т.б. зерттеушілер жиілік сөздіктер құрастыру ісімен айналыса бастады. «Абай тілі сөздігінің» жиілік сөздігі, М.Әуезовтің 20 томдық шығармалар тексінің жиілік сөздіктері, сонымен қатар, ертегі тілінің, публицистика тілінің, математика терминдерінің жиілік сөздіктері жарық көрді.

Осы жиілік сөздіктерден кейін қазақ тілінде жиілік сөздіктер құрастыру ісі тоқтап қалды. 2012 – 2014 жылдары А.Байтұрсынұлы атындағы ТБИ Қолданбалы лингвистика бөлімі аз көлемдегі қаржымен «Қазақ тілінің квантитативтік құрылымы» (Қазақ тілінің жиілік сөздігі) деген гранттық тақырып аясында бұрын жарық көрген сөздіктердің басын біріктіру арқылы әртүрлі стильдерді қамтитын біртұтас жиілік сөздікті құрастырып, баспаға ұсынды.

2016 жылы Білім және ғылым министрі Сағадиевтің тапсыруымен мемлекеттік тілді оқытудың тиімді әдістемесін жасау мақсатында лексика-грамматикалық минимумдар үшін қажетті қазақ тілінің жиілік сөздігін құрастыру ісі А.Байтұрсынұлы атындағы Тіл білімі институтына тапсырылған болатын. Осы тапсырма бойынша қысқа мерзімнің ішінде «Жалпы білім берудегі қазақ тілінің жиілік сөздігі» жарық көрді.

«Жалпы білім берудегі қазақ тілінің жиілік сөздігін» құрастыруға көркем әдебиет, публицистикалық стиль, ғылыми, публицистикалық, сөйлеу стильдер мәтіндерін қамтылған, яғни қазақ тілінің 5 функционалдық стилі сөздіктің базасын құрайды. 36 265 сөзтізбеден (сөз, лексикалық бірлік) тұратын бұл сөздік дәстүрлі сөздіктер түрлері болып табылатын үш түрлі құрылымда (Әліпбилі жиілік сөздік; Кері-Әліпбилі жиілік сөздік; Жиілікті әліпбилі сөздік) жасалған. Сонымен қатар бұл сөздік әрбір стильдер бойынша да «Таралым сөздігінде» (Распределительный словарь) жіктеліп берілген.

Жиілік сөздік жасаудың әдіс-тәсілдері әртүрлі. Соның бірі – мәтінді сөзформалар сөздігіне түсіріп, сөзді түбірге келтіру. Екіншісі – мәтінді морфологиялық анализатор арқылы өткізіп, түбір мен оған қойылатын сөз табы белгісімен бірге автоматты жолмен алу. Біріншісі көп қол жұмысын керек етеді. Мұндайда қосымша жалаңғанда түбір сөздердің омонимдерін ажырату мүмкін болса, түбір формадағы сөздерге келгенде олардың сөз табына қатысын анықтау қиын болады.

Екіншісі оңай, тез жүзеге асырылғанмен, жүз пайыздық шынайылықты бермейді. Бұл әдісте те омонимдер автоматты түрде ажыратылмайды. Автоматты жолмен омонимдері ажыратылған жиілік сөздіктерді тек омоним ажыратылған мәтіндерден ғана алуға болады.

Осы орайда Қолданбалы лингвистика бөлімі қызметкерлері проза жанры (М.Әуезов, Ә.Кекілбаев т.б. ) бойынша 500 000 сөзқолданыстан тұратын мәтіндерге «қолдап» морфологиялық белгіленім қойған болатын. Біз ұсынып отырған жиілік сөздік осы сөз таптары қолмен қойылған, омонимдері ажыратылған, яғни омоним жоқ мәтіннен алынды. Сөздің жиілік сөздігі бойынша әртүрлі сипаттағы жиілік талдаулар жасалды. Мәселен, 500 000 сөзқолданыстан тұратын мәтінде Қ және Т әрпінен басталатын сөз саны 1019 реестрлік сөзді құрап, барлық реестрлік сөздер ұзындығының **22,539**



## ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР

1 Бектаев К.Б., Ризаев С. О критерии принадлежности диалектов к одному языку // Статистика казахского текста. – Алматы, 1973. – С.643-685.

2 Исенгельдина А. Факторы, определяющие относительную частотность фонем// Статистика казахского текста. – Алматы, 1973. – С.643-685.

3 Жұбанова Е.Б. Қ.Жұбанов және қазақ тіліндегі буын жігін табу теориясы // Қазақ тілінің теориялық негізін қалаған лингвист, қоғам қайраткері, профессор Қ.Қ.Жұбановтың 110 жылдық және белгілі лингвист ғалым, профессор Е.Қ.Жұбановтың 80 жылдық мерейтойларына арналған «Жұбанов тағылымы» VII халықаралық ғылыми конференция материалдары. – Ақтөбе: 4-5 желтоқсан 2009 жыл. – 168-172 бб.

4 Джубанов А.Х. К вопросу о графемной статистике казахского текста // Вопросы казахской фонетики фонологии. – Алма-Ата: Наука, 1979. –С.49-52.

5 Жұбанов А., Жаңабекова А. Қазақ жазба мәтіндеріндегі әріп қолданысының статистикасы //Қазақстанның ғылымы мен өмірі/халықаралық ғылыми-көпшілік журнал. – Астана, 2017. – № 5. – 231-235 бб.

6 Шарипбай А. Проблемы перевода казахской письменности на латинский алфавит. Астана, 2017. – 138 с.

7 Қазақ тілінің сөзформалар жиілік сөздігі. – Алматы, 2016. (электрондық).

8 Жалпы білім берудегі қазақ тілінің жиілік сөздігі. – Алматы, 2016. – 1472 б.

9 Карбозова Б. Латын қарпіне негізделген түрік пернетақтасының тәжірибесі және қазақ пернетақтасына әріп-түйемішкітерді орналастыру реті туралы //Жаңа ұлттық әліпби негізінде қазақ жазуын реформалау теориясы мен практикасы. – Алматы: Қазақ тілі, 2016. – 560 б.

10 Агманов Е., Жұбанов А. Распределения частот появления сочетаний знаков в орхоно-енисейской письменности //Статистика казахского текста. – Алматы, 1973. – С.630-634.

11 Бектаев Қ. Статистика фонемосочетаний в каракалпакском языке // Статистика казахского текста. – Алматы, 1973. – С.635-642.

12 Джубанов А.Х. Статистика буквосочетаний в казахских текстах //Статистика казахского текста. – Алматы, 1973. – С.614-629.

13 Байтұрсынов А. Тіл тағылымы. – Алматы: Ана тілі, 1992.

14 Жұбанов Қ. Қазақ тілі бойынша зерттеулер. – Алматы: Ғылым, 1966.

15 Кеңесбаев І., Мұсабаев Ғ. Қазіргі қазақ тілі. – Алматы : Ғылым, 1975. – 302 б.

16 Мырзабеков С. Қазақ тілі фонетикасы. – Алматы: Қазақ университеті, 2004. – 247 б.

- 17 Джунибеков А. Сингармонизм в казахском языке. – Алма-Ата: Наука, 1985. –97 с.
- 18 Касевич В.Б. Слог в общем и дальневосточном языкознании // Теория языка. Методы его исследования и преподавания. – Л.: Наука, 1981. – С. 141-146.
- 19 Бектаев Қ. Қазақ тілі буындарының алфавитті-жиілік сөздігі //Статистика казахского текста. – Алматы, 1973. – С.566-610.
- 20 Ахабаев Ә., Бектаев Қ. Зат есім сөздердің морфологиялық құрылымына статистикалық талдау //Статистика казахского текста. – Алматы, 1973. – С.49-87.
- 21 Ахабаев Ә. Есімдіктердің морфологиялық құрылымына статистикалық талдау //Статистика казахского текста. – Алматы, 1973. – С.88-116.
- 22 Мырзабеков С.Туынды түбір егістіктердің құрылымына статистикалық талдау //Статистика казахского текста. – Алматы, 1973. – С.117-135.
- 23 Белботаев А. Ғылыми-техникалық стильдегі сын есім аффикстерінің қолданылуы туралы //Статистика казахского текста. – Алматы, 1973. – С.225-230.
- 24 Джубанов А.Х., Жубанов Е.Х. Статистика лексико-морфологических форм прилагательных в романе М.Ауезова «Абай жолы» («Путь Абая») //Статистика казахского текста. – Алматы, 1973. – С.231-248.
- 25 Жұбанов А.Қ. Основные принципы формализации содержания казахского текста. – Алматы, 2002.
- 26 Қазақ тілінің мәтіндер корпусы //Қолданбалы лингвистика бөліміндегі серверде 89.250.84.132
- 27 Жұбанов А., Жаңабекова А., Карбозова Б., Қожахметова А. Қазақ тілінің жиілік сөздігі. – Алматы: «Қазақ тілі» баспасы, 2016. – 665 б.
- 28 Фрумкина Р.М. Статистические методы изучения лексики. – М., 1964. – 115 с.

## МАЗМҰНЫ

<b>АЛҒЫ СӨЗ</b> .....	<b>3</b>
<b>КІРІСПЕ</b> .....	<b>5</b>
<b>1 ҚАЗАҚ ӘРІПТЕРІ МЕН ӘРІП ТІРКЕСТЕРІНІҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ</b> .....	<b>8</b>
1.1 Қазақ әріптері мен әріп тіркестерінің жиілік сөздігін құрастырудың қажеттілік негіздемесі .....	8
1.2 Қазақ әріптерінің жиілік сөздігін құрастыру тарихы .....	8
1.3 Қазақ әріптерінің жиілік сипаттамасы .....	18
1.4 Қазақ тіліндегі әріп тіркестерінің жиілік сөздігін құрастыру тарихы .....	23
1.5 Әріп тіркестерінің жиілік сипаттамасы .....	25
1.6 Ұлттық қазақ әліпбиін пернетақтада орналастырудың статистикалық негіздері .....	39
<b>2 ҚАЗАҚ ТІЛІНДЕГІ БУЫНДАРДЫҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ</b> .....	<b>45</b>
2.1 Қазақ тіліндегі буындардың жиілік сөздігін құрастырудың қажеттілік негіздемесі .....	45
2.2 Қазақ тіліндегі буындардың жиілік сөздігінің тарихы .....	48
2.4 Қазақ тілі буындарының статистикалық сипаттамасы .....	51
<b>3 ҚАЗАҚ ТІЛІНДЕГІ ҚОСЫМШАЛАРДЫҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ</b> .....	<b>73</b>
3.1 Қазақ тіліндегі түрленім қосымшалар мен түрленім формалар жиілік сөздігін құрастырудың қажеттілік негіздемесі .....	73
3.2 Қазақ тіліндегі морфемалар (сөз таптары, қосымшалар) жиілік сөздігінің тарихы .....	73
3.3 Қазақ тіліндегі түрленім қосымшаларының статистикалық сипаттамасы .....	75
3.4 Қазақ тіліндегі түрленім формалардың статистикалық сипаттамасы .....	100
<b>4 ҚАЗАҚ ТІЛІНДЕГІ СӨЗДЕРДІҢ ЖИЛІК СӨЗДІГІНІҢ СИПАТТАМАСЫ</b> .....	<b>117</b>
4.1 Қазақ тіліндегі сөздердің жиілік сөздігін құрастырудың қажеттілік негіздемесі .....	117
4.2 Қазақ тіліндегі сөздердің жиілік сөздігін құрастыру тарихы .....	118
4.3 Қазақ тіліндегі сөздердің жиілік сөздігінің сипаттамасы .....	121

4.3.1 Қазақ тіліндегі сөздің әліпбилі жиілік сөздігі .....	122
4.3.2 Қазақ тіліндегі сөздердің кері-әліпбилі жиілік сөздігі .....	131
4.3.3 Қазақ тіліндегі сөздердің жиілікті әліпбилі жиілік сөздігіне қатысты статистикалық сипаттама .....	138
<b>ҚОРЫТЫНДЫ.....</b>	<b>149</b>
<b>ПАЙДАЛАНЫЛҒАН ӘДЕБИЕТТЕР .....</b>	<b>156</b>

**А.Қ. ЖҰБАНОВ, А.Ә.ЖАҢАБЕКОВА**

**ҚАЗАҚ ЖАЗУЫН ЛАТЫН ҚАРПІНЕ  
КӨШІРУДІҢ СТАТИСТИКАЛЫҚ  
НЕГІЗДЕРІ**

Тех. редакторы Б. Қосдәулетов

Теруге 00. 11. 2018 берілді.  
Басуға 00. 11. 2018 қол қойылды.  
Пішімі 60x84 1/16. Офсеттік басылым.  
Қаріп түрі «Times New Roman»  
Шартты баспа табағы 9,3.  
Таралымы 1000