

**«ЛАТЫН ӘЛІПБИІ – ӘЛЕМ ҚАЗАҚТАРЫНЫҢ РУХАНИ
ТҰТАСТЫҒЫНЫҢ НЕГІЗІ»
РЕСПУБЛИКАЛЫҚ ҒЫЛЫМИ-ПРАКТИКАЛЫҚ
КОНФЕРЕНЦИЯ МАТЕРИАЛДАРЫ**

**Қазақ тілі әріптері мен әріп тіркестерінің жиіліктерін анықтау
латын әліпбиіне көшуде өз септігін тигізеді**

Жұбанов А.Қ.

филология ғылымдарының докторы, профессор

Жаңабекова А.Ә.

филология ғылымдарының докторы, профессор

Тоқмырзаев Д.О.

инженер-программист

Ключевые слова: *Частотный словарь, слова, словоформы, буквы, сочетания букв, графемы, сочетания графем, абсолютная частота, относительная частота, высокочастотные и низкочастотные буквы и их сочетания, процент покрываемости текста, клавиатура, расположение букв на клавиатуре, позиционное расположение букв, статистика букв и буквосочетаний, статистика слогов.*

Резюме: *В статье рассматриваются результаты статистических исследований букв и буквосочетаний, а также слогов казахского языка, с учетом их позиционных положений в словоформах. Данное статистическое исследование проведено с целью поиска удобного расположения букв казахского алфавита на клавиатуре компьютера.*

Кілт сөздер: *Жиілік сөздік, сөз, сөзформа, буын, әріп, әріп тіркесі, екіәріптік тіркес, жиілік, абсолютті жиілік, қатынастық жиілік, жоғары жиілікті, төмен жиілікті әріптер және әріп тіркестері, клавиатура, қазақ әліпбиінің клавиатурада ұтымды орналасуы және т.б.*

Түйіндеме: *Мақалада қазақ тілінің жазба мәтінін әріптік (графемалық) деңгейде және қазақ тілі сөзформаларының құрылымына, яғни әріптерге, әріп тіркестеріне, буындарға статистикалық әдіспен зерттеу нәтижелері баяндалады. Сөз құрылымындағы әріп пен буынды табиғи тілдің жазба мүмкіндігі арқылы бейнелейтін ең бір қарапайым түрдегі көрнекі тілдік бірліктері ретінде санау негізгі ұстаным болып табылады. Графемдік деңгейдегі статистикалық зерттеудің негізгі мақсаты – компьютер пернетақтасында қазақ әріптерін ұтымды орналастыру.*

2016 жылы А.Байтұрсынұлы атындағы Тіл білімі институты Қазақстан Республикасы Білім және ғылым министрі Е.Сағадиевтің тапсырмасы бойынша «Қазақ тілінің ұлттық корпусын әзірлеу және жасау» зерттеу жобасы аясында «Жалпы білім берудегі қазақ тілінің жиілік сөздігін» шығарды. Бұл сөздікте 7 миллионнан астам сөзқолданыстан тұратын қазақ тілінің 5 стилінен мәтіндер қамтылып, 61 түрленім нұсқасы дайындалды. Сөзтізбе саны – 36262 сөз.

Аталған 7 миллионнан астам сөзқолданыстан тұратын қазақ тілінің мәтінінен «Қазақ тілінің сөзформалар жиілік сөздігін» де түзілетінін ескерсек, бұл сөздікті де қазақ графемаларының мәтіндерде кездесу мүмкіндіктерін толық қамтыған жиілік сөздік деп те санауға болады.

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 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 әріп тіркестерін мысалмен дәйектейтін сөздерді қазақ тілінің мәтіндер корпусынан іздестіріп, көптеген мысал сөздерді табу сияқты мәселелерді шешуде аса қажет.

Төмендегі **1.1-кестеде** Қазақ тіліндегі әріптердің жиілікті-әліпбилі жиілік сөздігі, яғни әріптердің қайталану жиіліктерінің ең көп кездесуінен бастап, біртіндеп кему тәртібімен орналасқан әріптер тізімі көрініс тапты:

1.1-кестедегі мәліметтерді 3 топқа бөліп қарастыратын болсақ, 1-топқа ең жиі қолданатын 19 әріптерді: *а, е, ы, н, і, т, р, л, д, с, м, қ, о, к, г, б, й, у, з* жатқызуға болады. Бұл әріптердің ішінде ең жиі қолданатыныс тауып, 1-орынға ие «*а*» әріпі **8.630862** рет қолданып, мәтіннің **12,796** пайызын қамтиды екен, ал жиілік жағынан 2-орынға ие болатын «*е*» әріпі **5.660025** рет қолданып, мәтіннің **8.391** пайызын қамтиды. Енді осы «*а*», «*е*» әріптерінің бірігіп қолданысы барлық мәтіннің **21,187** пайызын қамтып жатқанына **1.1-кесте** бойынша көз жеткізуге болады. Сол сияқты, 1-топқа жататын ең жиі қолданатын 19 әріп (*а, е, ы, н, і, т, р, л, д, с, м, қ, о, к, г, б, й, у, з*) барлық мәтіннің **87,630** пайызын қамтып жатқанын айтуға болады.

1.1-кесте. Қазақ тіліндегі әріптердің жиілікті-әліпбилі жиілік сөздігі

№	Әріп	Абсолютті жиілігі	Бір топ әріптің мәтінді қамту пайызы	№	Әріп	Абсолютті жиілігі	Бір топ әріптің мәтінді қамту пайызы
1	2	3	4	1	2	3	4
1	а	8630862	12,796	22	г	903479	91,763
2	е	5660025	21,187	23	ң	872812	93,057
3	ы	5378832	29,161	24	и	866761	94,342
4	н	4192570	35,377	25	ж	738252	95,437
5	і	4154893	41,537	26	ұ	673484	96,435

6	т	4066134	47,565	27	ө	655787	97,408
7	р	4028709	53,537	28	ү	598755	98,295
8	л	3824239	59,207	29	ә	524073	99,072
9	д	3413312	64,267	30	я	231433	99,415
10	с	2708334	68,283	31	х	105688	99,572
11	м	2040371	71,308	32	ц	81165	99,692
12	қ	1876617	74,090	33	ф	58378	99,779
13	о	1639224	76,520	34	в	55978	99,862
14	к	1633360	78,941	35	э	30250	99,907
15	ғ	1256146	80,804	36	ь	25076	99,944
16	б	1229938	82,627	37	ю	20877	99,975
17	й	1175543	84,370	38	h	5406	99,983
18	у	1103283	86,006	39	ъ	5148	99,991
19	з	1095869	87,630	40	ч	3702	99,996
20	п	978676	89,081	41	ш	1679	99,998
21	ш	905698	90,424	42	ё	1018	100

Енді 2-топқа, қолдану жиілігі жағынан абсолютті жиіліктері **978676-524073** аралығында жататын сан жағынан 10-ға тең (*n, ш, з, ң, и, ж, ұ, ө, ү, ә*) әріптер екен. Аталған 10 әріп қайталана қолдана келіп, барлық мәтіннің 36,443 пайызын қамтып жатыр. Сол сияқты, 3-топқа сирек қолданатын басқа тілдерден енген келесі 13 әріпті жатқыздық: *я, х, ц, ф, в, э, ь, ю, h, ъ, ч, ш, ё*.

3-топтағы кірме дыбыстарға жататын бұл әріптер 7 миллионнан астам сөзқолданыстан тұратын қазақ тілінің 5 стилі бойынша алынған мәтіндердің бар болғаны 0,928 пайызын ғана қамтиды екен, яғни ол 1 пайызға да жетпейтін мәтінді қамту пайызы.

Жоғарыда сөз болған қазақ әріптерінің статистикалық мәліметтері қазақтың ұлттық компьютерлік пернетақсына қазақ әліпбиін латынға көшіру кезінде пайдалануға мүмкіндік береді.

Енді қазақ әріптерінің екіәріптік тіркестерінің статистикалық сипаттамасына қысқаша тоқталайық.

1.2-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
8	<i>ер</i>	481450	13,173909

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
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

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
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
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

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
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
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>ығ</i>	89655	73,515528
108	<i>зі</i>	89577	73,780530

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
109	<i>уы</i>	89453	74,045166
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>пт</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>жсы</i>	59382	80,918179
142	<i>кү</i>	59169	81,093223
143	<i>жо</i>	58694	81,266862
144	<i>тұ</i>	57334	81,436478
145	<i>еу</i>	54592	81,597981
146	<i>ни</i>	54403	81,758926
147	<i>жә</i>	53682	81,917737
148	<i>ка</i>	53591	82,076280
149	<i>ңд</i>	53060	82,233251
150	<i>жү</i>	52798	82,389447
151	<i>аш</i>	52156	82,543744
152	<i>лм</i>	51437	82,695914
153	<i>от</i>	50604	82,845620
154	<i>ти</i>	50538	82,995130

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
155	<i>лг</i>	49924	83,142824
156	<i>өл</i>	49481	83,289208
157	<i>ез</i>	48885	83,433828
158	<i>лу</i>	48642	83,577729
159	<i>үй</i>	47945	83,719568
160	<i>ха</i>	47869	83,861183
161	<i>іг</i>	47318	84,001167
162	<i>сө</i>	45850	84,136809
163	<i>уа</i>	45677	84,271939
164	<i>сқ</i>	45262	84,405841
165	<i>тк</i>	44909	84,538698
166	<i>ик</i>	44264	84,669648
167	<i>бы</i>	44121	84,800174
168	<i>рс</i>	43707	84,929476
169	<i>үш</i>	42635	85,055606
170	<i>әс</i>	42547	85,181476
171	<i>үс</i>	42402	85,306917
172	<i>йі</i>	40912	85,427950
173	<i>ни</i>	40860	85,548830
174	<i>зы</i>	40487	85,668605
175	<i>ур</i>	40401	85,788127
176	<i>ин</i>	40271	85,907264
177	<i>лт</i>	40245	86,026323
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>өт</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>үг</i>	35743	87,834024
194	<i>әл</i>	35728	87,939721
195	<i>ми</i>	35603	88,045048
196	<i>ұл</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>жү</i>	33023	88,548196

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
201	ңі	32705	88,644950
202	оғ	32191	88,740183
203	ял	32129	88,835233
204	нт	31942	88,929729
205	ұм	31057	89,021608
206	ци	30876	89,112950
207	үй	30113	89,202036
208	іш	29962	89,290675
209	өп	29755	89,378701

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
210	ли	29713	89,466603
211	ұл	29290	89,553254
212	ск	29049	89,639192
213	мұ	29004	89,724997
214	ая	28939	89,810609
215	лс	28186	89,893994
216	нг	27874	89,976456
217	ңы	27579	90,058045

Қазақ тілінің екіәріптік тіркестерінің жиілік сөздігін құрастыру үшін жоғарыда аталған 7 миллионнан астам сөзқолданыстан тұратын қазақ тілінің мәтінінен құрастырылған «Қазақ тілінің сөзформалар жиілік сөздігі» негіз болды.

Компьютер көмегімен құрастырылған екіәріптік жиілік сөздікте әртүрлі тіркестердің реестрлік рет саны 1160-қа тең, ал осы екіәріптік тіркестер қайталанып қолдана 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 қамтытын жоғары жиілікті зонаға жататын 217 екіәріптік тіркестер бөлігін **1.2-кесте** бойынша қарастыруға болады.

Қазақ тілінің ұлттық компьютерлік пернетақтасында **1.1-кестедегі** жиі кездесетін 1-топтағы 19 әріп екі бөлікке бөлініп, пернетақтаның ортаңғы бөлігіне орналасады. Сол сияқты, **1.2-кестедегі** екіәріптік тіркестердің ең жиі қолданатын, мәселен мәтіннің 90 пайызын қамтитын рет сандары 1-217 аралығын қамтитын тіркестердің екінші әріптерін **1.1-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, яғни жиі қолдана отырып, мәтіннің 90 пайызын қамтитыны жайлы **1.3-кесте** және **1.4-кесте** арқылы танысуға болады.

1.3-кесте бойынша сөз басындағы екіәріптік тіркестердің статистикалық сипатын айқындауға болады.

**1.3-кесте. ҚАЗАҚ ТІЛІНІҢ СӨЗ БАСЫНДАҒЫ ЕКІӘРШПТІК
ТІРКЕСТЕРІНІҢ
ЖИЛІК СӨЗДІГІ (ҮЗІНДІ)**

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы	Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
1	2	3	4	1	2	3	4
1	<i>қа</i>	314514	5,33657	42	<i>не</i>	36931	62,32549
2	<i>жа</i>	228241	9,20929	43	<i>от</i>	36633	62,94707
3	<i>ба</i>	222845	12,99045	44	<i>ау</i>	36611	63,56827
4	<i>ке</i>	177014	15,99396	45	<i>оң</i>	36401	64,18591
5	<i>бі</i>	149781	18,53540	46	<i>ре</i>	36121	64,79880
6	<i>бо</i>	144030	20,97925	47	<i>мә</i>	35588	65,40265
7	<i>та</i>	142349	23,39458	48	<i>ос</i>	34590	65,98956
8	<i>кө</i>	126340	25,53828	49	<i>да</i>	34176	66,56945
9	<i>ме</i>	123027	27,62576	50	<i>жұ</i>	32232	67,11635
10	<i>са</i>	116450	29,60164	51	<i>кі</i>	31591	67,65238
11	<i>де</i>	116241	31,57398	52	<i>ес</i>	31011	68,17856
12	<i>қо</i>	108225	33,41031	53	<i>оқ</i>	30805	68,70125
13	<i>ал</i>	99351	35,09607	54	<i>ту</i>	30296	69,21530
14	<i>бе</i>	91785	36,65344	55	<i>өт</i>	28845	69,70473
15	<i>со</i>	81582	38,03770	56	<i>ер</i>	28770	70,19289
16	<i>же</i>	75181	39,31335	57	<i>па</i>	28510	70,67664
17	<i>құ</i>	73538	40,56111	58	<i>ад</i>	27917	71,15033
18	<i>тү</i>	70582	41,75873	59	<i>ше</i>	27320	71,61389
19	<i>ай</i>	65318	42,86702	60	<i>ко</i>	27211	72,07559
20	<i>бұ</i>	59954	43,88430	61	<i>ақ</i>	27077	72,53503
21	<i>ор</i>	58427	44,87567	62	<i>үш</i>	25670	72,97059
22	<i>қы</i>	58085	45,86124	63	<i>сы</i>	25579	73,40460
23	<i>кү</i>	57775	46,84154	64	<i>мұ</i>	25496	73,83721
24	<i>жо</i>	57314	47,81403	65	<i>ас</i>	25250	74,26564
25	<i>тұ</i>	55624	48,75784	66	<i>тө</i>	24990	74,68967
26	<i>ма</i>	55004	49,69113	67	<i>ет</i>	24806	75,11057
27	<i>ар</i>	54779	50,62060	68	<i>ол</i>	24071	75,51899
28	<i>жә</i>	53475	51,52795	69	<i>за</i>	24004	75,92629
29	<i>шы</i>	51825	52,40730	70	<i>тә</i>	23876	76,33141
30	<i>өз</i>	50541	53,26486	71	<i>ға</i>	23492	76,73001
31	<i>те</i>	50297	54,11828	72	<i>мы</i>	23137	77,12259
32	<i>жү</i>	49984	54,96639	73	<i>ой</i>	22516	77,50464
33	<i>ат</i>	49131	55,80003	74	<i>әр</i>	22221	77,88167
34	<i>ті</i>	48029	56,61497	75	<i>ел</i>	21756	78,25082
35	<i>жы</i>	47056	57,41340	76	<i>ұл</i>	20966	78,60657
36	<i>ек</i>	45959	58,19322	77	<i>әл</i>	20143	78,94835
37	<i>сө</i>	43937	58,93873	78	<i>іс</i>	19666	79,28203
38	<i>ша</i>	43897	59,68356	79	<i>бө</i>	19398	79,61117
39	<i>се</i>	42067	60,39734	80	<i>ем</i>	18988	79,93335
40	<i>то</i>	38897	61,05733	81	<i>ед</i>	18731	80,25118
41	<i>ха</i>	37809	61,69886	82	<i>мү</i>	18613	80,56700

1	2	3	4	1	2	3	4
83	<i>сұ</i>	18011	80,87260	104	<i>сә</i>	12973	86,43772
84	<i>дә</i>	17976	81,17761	105	<i>бұ</i>	12953	86,65751
85	<i>үй</i>	17527	81,47500	106	<i>бә</i>	12880	86,87605
86	<i>ан</i>	17500	81,77194	107	<i>жи</i>	12741	87,09224
87	<i>би</i>	17496	82,06880	108	<i>ки</i>	12615	87,30628
88	<i>на</i>	17402	82,36407	109	<i>мо</i>	12530	87,51889
89	<i>ен</i>	17090	82,65405	110	<i>аш</i>	12060	87,72352
90	<i>әд</i>	16843	82,93984	111	<i>жі</i>	12013	87,92735
91	<i>ты</i>	16671	83,22271	112	<i>мі</i>	11877	88,12887
92	<i>су</i>	16160	83,49690	113	<i>ұй</i>	11858	88,33008
93	<i>өл</i>	15674	83,76286	114	<i>ағ</i>	11826	88,53074
94	<i>өн</i>	15137	84,01970	115	<i>аз</i>	11363	88,72354
95	<i>си</i>	15010	84,27438	116	<i>ең</i>	10523	88,90209
96	<i>іш</i>	14993	84,52878	117	<i>ая</i>	10468	89,07971
97	<i>ұл</i>	14866	84,78102	118	<i>сұ</i>	10463	89,25724
98	<i>ұс</i>	14847	85,03294	119	<i>дү</i>	10427	89,43416
99	<i>пр</i>	14819	85,28438	120	<i>ин</i>	10291	89,60878
100	<i>кә</i>	14494	85,53031	121	<i>әс</i>	9983	89,77817
101	<i>әк</i>	13795	85,76438	122	<i>до</i>	9936	89,94676
102	<i>ти</i>	13599	85,99512	123	<i>жө</i>	9872	90,11426
103	<i>өм</i>	13112	86,21760				

Мысалы, ең жиі қолданатын екіәріптік тіркестердің алғашқы әріпіне қарай олардың екінші әріптерінің қандай екенін анықтауға болады. **1.3-кестеде** рет саны 1-ші болып тұрған «**қа**» екіәріптік тіркес мәтінде ең жиі кездесіп, оның абсолютті жиілігі 314514-ке тең және барлық мәтінді қамту пайызы – 5,33%. Енді, мысал үшін, осы «**қ**» әріпінің мәтіннің 90 пайызын қамтып жатқан 123 тіркестер ішінде қандай әріптермен тіркесіп жиі қолданатынын анықтайық. Мысалы, жиілік сөздіктің жиі қолданған аясында орналасқан **қа, қо, құ, қы** тіркестерін, сол сияқты «**ж**» әріпінің 11 түрлі тіркестері: **жа, же, жо, жә, жү, жы, жы, жұ, жи, жі, жө** жиі қолданып кездескен екен. Сол сияқты, «**т**» әріпінің 8 түрлі жиі тіркесін келтірсек, олар: **та, тү, тұ, те, то, ту, ты, ти**. Міне мұндай мысалдарды қазақ әліпбинің басқа да әріптері бойынша статистикалық мәліметтер мақсатында келтіруге болады.

Екіәріптік тіркестердің жиілік сөздіктерінің үзінділері **1.2, 1.3, 1.4-кестелерде** көрініс тапты, оларда тіркестердің сөз ішінде орналасу орнына қарай бөліп қарастырылды. Мысалы, **1.2-кестеде** әріп тіркесі сөзформа ішінде алатын орнына қатыссыз, **1.3-кестеде** – сөз басындағы тіркестер, **1.4-кестеде** – сөз соңындағы тіркестердің статистикасы жайлы мәліметтер қамтылды. Енді оларды қазақ тілінің компьютерлік ұлттық пернетақтасын жасауда қалай пайдаланамыз деген сұрақ туындауы мүмкін.

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

Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы	Рет саны №	Әріп тіркесі	Абсолютті жиілік	Бір топ әріп тіркестерінің мәтінді қамту пайызы
1	2	3	4	1	2	3	4
1	<i>ен</i>	319537	5,41253	37	<i>ек</i>	46525	72,87324
2	<i>ан</i>	272130	10,02205	38	<i>ім</i>	44391	73,62517
3	<i>ың</i>	240700	14,09919	39	<i>ау</i>	43245	74,35768
4	<i>ды</i>	240100	18,16617	40	<i>қа</i>	43056	75,08699
5	<i>ын</i>	195229	21,47309	41	<i>ыс</i>	42655	75,80951
6	<i>ді</i>	187867	24,65530	42	<i>ым</i>	39675	76,48155
7	<i>ін</i>	176123	27,63859	43	<i>ша</i>	37192	77,11154
8	<i>ық</i>	165604	30,44370	44	<i>ет</i>	36345	77,72717
9	<i>ің</i>	157597	33,11319	45	<i>лі</i>	35812	78,33378
10	<i>ар</i>	149557	35,64648	46	<i>та</i>	35423	78,93380
11	<i>ып</i>	146130	38,12173	47	<i>ұл</i>	32593	79,48588
12	<i>да</i>	130213	40,32737	48	<i>сі</i>	32030	80,02843
13	<i>ай</i>	113729	42,25378	49	<i>ру</i>	31677	80,56499
14	<i>не</i>	109065	44,10120	50	<i>ес</i>	30999	81,09007
15	<i>де</i>	102506	45,83752	51	<i>ні</i>	30949	81,61431
16	<i>га</i>	101192	47,55157	52	<i>ке</i>	30465	82,13035
17	<i>ты</i>	98951	49,22767	53	<i>са</i>	29680	82,63309
18	<i>ік</i>	95044	50,83759	54	<i>ол</i>	27248	83,09463
19	<i>ер</i>	91362	52,38514	55	<i>ші</i>	26643	83,54593
20	<i>на</i>	89806	53,90634	56	<i>ей</i>	26551	83,99567
21	<i>сы</i>	89508	55,42248	57	<i>се</i>	25855	84,43362
22	<i>ғы</i>	80934	56,79340	58	<i>ыр</i>	24862	84,85474
23	<i>ті</i>	80568	58,15811	59	<i>ат</i>	24729	85,27362
24	<i>ге</i>	80523	59,52207	60	<i>ра</i>	24671	85,69152
25	<i>ін</i>	78056	60,84423	61	<i>ас</i>	23516	86,08985
26	<i>ақ</i>	75244	62,11876	62	<i>ам</i>	23501	86,48792
27	<i>гі</i>	69045	63,28829	63	<i>те</i>	23009	86,87766
28	<i>лы</i>	68298	64,44517	64	<i>ла</i>	22852	87,26475
29	<i>рі</i>	64818	65,54310	65	<i>ма</i>	22668	87,64871
30	<i>ір</i>	64774	66,64029	66	<i>шы</i>	22100	88,02306
31	<i>ап</i>	64127	67,72651	67	<i>кі</i>	22094	88,39730
32	<i>еп</i>	56991	68,69187	68	<i>ту</i>	21449	88,76062
33	<i>ыз</i>	53348	69,59551	69	<i>уы</i>	20591	89,10940
34	<i>ры</i>	51962	70,47568	70	<i>ше</i>	19712	89,44330
35	<i>із</i>	47858	71,28633	71	<i>ыл</i>	18768	89,76120
36	<i>ны</i>	47161	72,08517	72	<i>оқ</i>	17887	90,06418

Осы сауалға қысқаша жауап берсек, біздің ұстанымыз бойынша ең жиі қолданатын 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 әріпі, тіркестің бірінші әріпінің оң жағына жақын орналасуы қажет деп білеміз. Ондай дәрежеге бір емес, бірнеше әріп «үміткер» болуы мүмкін. Осындай жағдайда екіәріптік тіркестің позициялық жағыдайына да мән берілуі керек.

Келесі айтайын дегеніміз, қазақ әліпбиін латынға көшіру мәселесінде тек әріп және әріп тіркестерінің статистикасынан басқа қазақ сөзінің құрылымдық бірліктің алғашқысы ретінде «буындық» бірлікті жатқызуға болады. Сол себепті біз «Қазақ тілі буындарының жиілік сөздігін» компьютер көмегімен жасап, қазақ тілі буындарына статистикалық сипаттама беруді де мақсат етіп қойдық.

Қазақ тілінің сөзформаларындағы буын жігін автоматты түрде ажырату мәселесі әлі толық шешімін таппағандықтан бұл процесті жартылай қолмен жүзеге асыруға тура келді.

Төменде көрініс тапқан «Қазақ тілі буындарының жиілік сөздігі» негізінде қазақ тіліндегі буындардың реттік саны 1225 тең болса, олардың қайталана қоладануы негізінде буындардың қосынды жиілігі 124063 тең екендігі дәлелденді.

Қазақ тіліндегі буындарды дыбыстық жағынан қарастыратын болсақ, оларды 3 топқа бөліп қарастыруға болатыны мәлім: 1) Ашық буын; 2) Тұйық буын; 3) Бітеу буын. Жиілік сөздікте буын түрлері шартты белгілермен таңбаланды: *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 **2.1-кестеде** көрініс тапты. Жиілік сөздіктің реттік саны бойынша барлық *ашық буынға* (**АБ**) жататын бірліктер саны 173-ке тең, ал олар қайталана қолдана келіп мәтіннің 54,59 пайызын қамтитынына көз жеткізуге болады.

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

Сол сияқты, *бітеу буын* (**ББ**) жайлы сөз ететін болсақ, олар кері-әліпбилі жиілік сөздікте ашық буынға қарағанда сандық жағынан айтарлықтай жиі қолданады деуге болады. Мәселен, рет саны бойынша бітеу буынның реестрдегі саны 952-ге тең де, ал барлық буын бірліктерінің мәтін ішінде қайталана қолданып, қазақ тілінің 5 стилі бойынша алынған мәтіннің 41 пайызын қамтиды екен. Қазақ тіліндегі бітеу буындардың кері-әліпбилі жиілік сөздігінің үзіндісі **2.2-кестеде** берілді.

2.1-кесте. Қазақ тіліндегі ашық буындардың кері-әліпбилі жиілік сөздігі

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы	Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
1	<i>aV/A</i>	1944	1,5669	42	<i>pa/C+V/A</i>	13	21,6172
2	<i>әV/A</i>	639	2,0820	43	<i>ca/C+V/A</i>	31	21,6422
3	<i>eV/A</i>	1083	2,9550	44	<i>ta/C+V/A</i>	100	21,7228
4	<i>uV/A</i>	50	2,9953	45	<i>ya/C+V/A</i>	60	21,7712
5	<i>oV/A</i>	1603	4,2873	46	<i>ша/C+V/A</i>	5	21,7752
6	<i>өV/A</i>	802	4,9338	47	<i>бе/C+V/A</i>	581	22,2435
7	<i>yV/A</i>	15	4,9459	48	<i>ge/C+V/A</i>	901	22,9698
8	<i>ұV/A</i>	403	5,2707	49	<i>де/C+V/A</i>	2226	24,7640
9	<i>үV/A</i>	375	5,5730	50	<i>же/C+V/A</i>	364	25,0574
10	<i>ыV/A</i>	30	5,5972	51	<i>ze/C+V/A</i>	97	25,1356
11	<i>iV/A</i>	138	5,7084	52	<i>ие/C+V/A</i>	35	25,1638
12	<i>ба/C+V/A</i>	997	6,5120	53	<i>йе/C+V/A</i>	243	25,3597
13	<i>га/C+V/A</i>	1	6,5128	54	<i>ке/C+V/A</i>	1366	26,4607
14	<i>ға/C+V/A</i>	1538	7,7525	55	<i>ле/C+V/A</i>	1058	27,3135
15	<i>да/C+V/A</i>	2522	9,7854	56	<i>ме/C+V/A</i>	705	27,8818
16	<i>жа/C+V/A</i>	1020	10,6075	57	<i>не/C+V/A</i>	1366	28,9829
17	<i>за/C+V/A</i>	256	10,8139	58	<i>ңе/C+V/A</i>	30	29,0070
18	<i>йа/C+V/A</i>	307	11,0613	59	<i>пе/C+V/A</i>	56	29,0522
19	<i>қа/C+V/A</i>	2520	13,0925	60	<i>ре/C+V/A</i>	454	29,4181
20	<i>ла/C+V/A</i>	1793	14,5378	61	<i>се/C+V/A</i>	638	29,9324
21	<i>ма/C+V/A</i>	876	15,2439	62	<i>те/C+V/A</i>	774	30,5562
22	<i>на/C+V/A</i>	1354	16,3352	63	<i>ye/C+V/A</i>	20	30,5724
23	<i>ңа/C+V/A</i>	107	16,4215	64	<i>ше/C+V/A</i>	436	30,9238
24	<i>па/C+V/A</i>	230	16,6069	65	<i>би/C+V/A</i>	142	31,0383
25	<i>ра/C+V/A</i>	1129	17,5169	66	<i>ги/C+V/A</i>	34	31,0657
26	<i>са/C+V/A</i>	1080	18,3874	67	<i>ди/C+V/A</i>	12	31,0753
27	<i>та/C+V/A</i>	1852	19,8802	68	<i>жи/C+V/A</i>	93	31,1503
28	<i>ya/C+V/A</i>	213	20,0519	69	<i>ки/C+V/A</i>	10	31,1584
29	<i>ха/C+V/A</i>	278	20,2760	70	<i>қи/C+V/A</i>	91	31,2317
30	<i>һа/C+V/A</i>	4	20,2792	71	<i>ли/C+V/A</i>	23	31,2503
31	<i>ша/C+V/A</i>	726	20,8644	72	<i>ми/C+V/A</i>	68	31,3051
32	<i>ба/C+V/A</i>	88	20,9353	73	<i>ни/C+V/A</i>	61	31,3542
33	<i>да/C+V/A</i>	70	20,9918	74	<i>пи/C+V/A</i>	6	31,3591
34	<i>жа/C+V/A</i>	357	21,2795	75	<i>ри/C+V/A</i>	118	31,4542
35	<i>за/C+V/A</i>	5	21,2835	76	<i>си/C+V/A</i>	73	31,5130
36	<i>иә/C+V/A</i>	6	21,2884	77	<i>ти/C+V/A</i>	48	31,5517
37	<i>кә/C+V/A</i>	103	21,3714	78	<i>ху/C+V/A</i>	10	31,5598
38	<i>лә/C+V/A</i>	1	21,3722	79	<i>ши/C+V/A</i>	16	31,5727
39	<i>мә/C+V/A</i>	216	21,5463	80	<i>бо/C+V/A</i>	579	32,0394
40	<i>нә/C+V/A</i>	37	21,5761	81	<i>до/C+V/A</i>	18	32,0539
41	<i>пә/C+V/A</i>	38	21,6068	82	<i>жо/C+V/A</i>	214	32,2264

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
83	<i>қо/С+V/A</i>	407	32,5544
84	<i>мо/С+V/A</i>	23	32,5730
85	<i>но/С+V/A</i>	1	32,5738
86	<i>ро/С+V/A</i>	6	32,5786
87	<i>со/С+V/A</i>	224	32,7592
88	<i>то/С+V/A</i>	103	32,8422

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
89	<i>шо/С+V/A</i>	17	32,8559
90	<i>бө/С+V/A</i>	103	32,9389
91	<i>гө/С+V/A</i>	16	32,9518
-----	-----	-----	-----

2.2-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+С+С/Б</i>	2	54,5908
175	<i>бүлк/С+V+С+С/Б</i>	5	54,5948
176	<i>сүлк/С+V+С+С/Б</i>	2	54,5965
177	<i>сылқ/С+V+С+С/Б</i>	3	54,5989
178	<i>даңқ/С+V+С+С/Б</i>	4	54,6021
179	<i>жарқ/С+V+С+С/Б</i>	2	54,6037
180	<i>жұрқ/С+V+С+С/Б</i>	1	54,6045
181	<i>шырқ/С+V+С+С/Б</i>	5	54,6085
182	<i>балл/С+V+С+С/Б</i>	2	54,6102
183	<i>тырл/С+V+С+С/Б</i>	1	54,6110
184	<i>дүрс/С+V+С+С/Б</i>	2	54,6126
185	<i>мырс/С+V+С+С/Б</i>	6	54,6174
186	<i>ғайт/С+V+С+С/Б</i>	1	54,6182
187	<i>жайт/С+V+С+С/Б</i>	11	54,6271
188	<i>қайт/С+V+С+С/Б</i>	12	54,6368
189	<i>тайт/С+V+С+С/Б</i>	1	54,6376
190	<i>жәйт/С+V+С+С/Б</i>	1	54,6384
191	<i>сөйт/С+V+С+С/Б</i>	2	54,6400
192	<i>жүйт/С+V+С+С/Б</i>	1	54,6408
193	<i>түйт/С+V+С+С/Б</i>	4	54,6440
194	<i>жалт/С+V+С+С/Б</i>	2	54,6456
195	<i>салт/С+V+С+С/Б</i>	12	54,6553
196	<i>шалт/С+V+С+С/Б</i>	1	54,6561
197	<i>зелт/С+V+С+С/Б</i>	1	54,6569
198	<i>селт/С+V+С+С/Б</i>	1	54,6577
199	<i>бұлт/С+V+С+С/Б</i>	3	54,6601
200	<i>жұлт/С+V+С+С/Б</i>	1	54,6609
201	<i>ғылт/С+V+С+С/Б</i>	3	54,6634

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
202	<i>кілт/С+V+С+С/Б</i>	4	54,6666
203	<i>лант/С+V+С+С/Б</i>	3	54,6690
204	<i>уант/С+V+С+С/Б</i>	1	54,6698
205	<i>зент/С+V+С+С/Б</i>	1	54,6706
206	<i>қарт/С+V+С+С/Б</i>	2	54,6722
207	<i>нарт/С+V+С+С/Б</i>	3	54,6746
208	<i>тарт/С+V+С+С/Б</i>	11	54,6835
209	<i>шарт/С+V+С+С/Б</i>	9	54,6908
210	<i>берт/С+V+С+С/Б</i>	1	54,6916
211	<i>дерт/С+V+С+С/Б</i>	4	54,6948
212	<i>зерт/С+V+С+С/Б</i>	28	54,7174
213	<i>керт/С+V+С+С/Б</i>	3	54,7198
214	<i>серт/С+V+С+С/Б</i>	2	54,7214
215	<i>жорт/С+V+С+С/Б</i>	1	54,7222
216	<i>төрт/С+V+С+С/Б</i>	17	54,7359
217	<i>жұрт/С+V+С+С/Б</i>	55	54,7802
218	<i>құрт/С+V+С+С/Б</i>	6	54,7851
219	<i>мұрт/С+V+С+С/Б</i>	2	54,7867
220	<i>күрт/С+V+С+С/Б</i>	4	54,7899
221	<i>түрт/С+V+С+С/Б</i>	3	54,7923
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>жырт/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

Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы	Рет саны №	Буын	Абсолютті жиілік	Бір топ буынның мәтінді қамту пайызы
230	<i>уырт/С+V+C+C/Б</i>	3	54,8181	247	<i>бөз/С+V+C/Б</i>	4	54,8592
231	<i>шырт/С+V+C+C/Б</i>	1	54,8189	248	<i>баз/С+V+C/Б</i>	87	54,9294
232	<i>гірт/С+V+C+C/Б</i>	1	54,8197	249	<i>даз/С+V+C/Б</i>	34	54,9568
233	<i>дірт/С+V+C+C/Б</i>	1	54,8205	250	<i>жаз/С+V+C/Б</i>	109	55,0446
234	<i>кірт/С+V+C+C/Б</i>	3	54,8230	---	-----	---	-----
235	<i>сірт/С+V+C+C/Б</i>	1	54,8238	1030	<i>быт/С+V+C/Б</i>	2	94,2279
236	<i>тірт/С+V+C+C/Б</i>	1	54,8246	1031	<i>зыт/С+V+C/Б</i>	35	94,2561
237	<i>қарш/С+V+C+C/Б</i>	4	54,8278	1032	<i>зыт/С+V+C/Б</i>	1	94,2570
238	<i>кірш/С+V+C+C/Б</i>	2	54,8294	1033	<i>йыт/С+V+C/Б</i>	4	94,2602
239	<i>жаб/С+V+C/Б</i>	4	54,8326	1034	<i>қыт/С+V+C/Б</i>	55	94,3045
240	<i>раб/С+V+C/Б</i>	20	54,8487	1035	<i>мыт/С+V+C/Б</i>	8	94,3110
241	<i>хаб/С+V+C/Б</i>	1	54,8496	1036	<i>ныт/С+V+C/Б</i>	11	94,3198
242	<i>шаб/С+V+C/Б</i>	1	54,8504	1037	<i>ңыт/С+V+C/Б</i>	1	94,3206
243	<i>гиб/С+V+C/Б</i>	1	54,8512	1038	<i>рыт/С+V+C/Б</i>	2	94,3222
244	<i>қоб/С+V+C/Б</i>	1	54,8520	1039	<i>уыт/С+V+C/Б</i>	4	94,3255
245	<i>сыб/С+V+C/Б</i>	4	54,8552	-----	-----	-----	-----
246	<i>мез/С+V+C/Б</i>	1	54,8560				

Қазақ тіліндегі тұйық буындардың кері-әліпбилі жиілік сөздігінің статистикалық мәліметтері бойынша барлық тұйық буындардың рет саны 100-ге тең және олар қайталана қолдана келіп барлық мәтіннің 4,44 пайызын ғана қамтиды екен. Ал енді олардың құрылымдық жағына келетін болсақ, оны екі түрге (вариантқа) бөліп қарастыруға болады. Біріншісі, 3 дыбыстан тұрады, оның алғашқысы дауысты (V), ал келесі екеуі дауыссыз дыбыстар (C/C), яғни моделі – V/C/C. Мұндай тұйық буынды бірліктердің жиілік сөздіктегі саны 9-ге ғана тең (*айт, әйт, өйт, ұйт, ұлт, ант, арт, ерт, өрт*) және олардың қайталана қолдану нәтижесі барлық мәтіннің 0,2 пайызын ғана қамтыйды екен. Осы 9 тұйық буындарының ішінде ең жоғары қолданатындар *айт* – 97 рет, *ұлт* – 67 рет және *өйт* – 24 рет қолданыс тапқан екен.

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

Ал қазір біз, ең жиі қолданылатын буын түрлеріне статистикалық сипаттама беру үшін «Қазақ тілі буындарының жиілікті-әліпбилі жиілік сөздігі» жайлы қысқаша сөз етпекпіз. Өздеріңізге белгілі, жиілікті-әліпбилі жиілік сөздіктің анықтамасы бойынша, бұл сөздікте буын бірліктерінің ең жиі қолданылатынан бастау алып, әрі қарай олардың біртіндеп кему

тәртібімен орналасатындығы жайлы айтылған. Сол себепті, жиілік сөздіктің бас жағында – буын түрлерінің ең жиі қолданылатын түрлері орналасқандықтан, сол буындардың статистикалық сипаттамасына қысқаша тоқталайық.

Жоғары жиілікті зонаға абсолютті жиіліктері 1000-нан жоғары, реттік саны 1–26 аралығындағы буындардың қайталана қолдана келіп, мәтіннің 30 пайызын қамтитын жиілік сөздік бөлігін жатқызуды ұйғардық және оны **бірінші топ** деп атауды ұйғардық. 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 қолданған ашық буындарды **екінші топқа** жатқызатын болсақ, олардың түрлерінің саны 73-ке тең және олар қайталана қолдану нәтижесінде барлық мәтіннің 22,45 пайызын қамтып жатыр.

Қорыта айтқанда, жазба тілдің сөйлеу тілі сияқты практикалық және теориялық тұрғыдан аса маңызды зерттеу нысаны екені мәлім. Мысалы, ақпаратты автоматты түрде өңдеу мен шартты белгілермен таңбалау (кодтау) кезеңінде сол тілдің графемаларының статистикалық сипаттамаларын білу аса маңызды. Графемаларды статистикалық әдіспен талдау арқылы алынған нәтижелер баспа ісін жетілдіруде және компьютерді қазақ тілін кирилден латынға ұтымды көшіру шараларында да аса қажет. Жазба тіліне тән ерекшеліктер, яғни жазба мәтіннің сандық және сапалық қатынастары қазақ тілінің әртүрлі әдеби жанрларынан орын алады.

Мақалада қазақ тілінің жазба мәтінін әріптік (графемалық) деңгейде және қазақ тілі сөзформаларының құрылымына статистикалық әдіспен зерттеуге әрекет жасалған. Себебі, әріп пен сөз құрылымындағы буын табиғи тілді жазба мүмкіндігі арқылы бейнелейтін ең бір қарапайым түрдегі көрнекі тілдік бірліктер болып саналады.

Зерттеу нәтижелері қазақ тілін фонологиялық және фонетикалық деңгейде тілдің морфемдік құрылымын талдау мен тілдік бірліктерді синтагматикалық тұрғыда қарастыру кезінде аса маңызды. Әріптердің, әріп тіркестерінің жалпы және шептік орналасу мүмкіндіктерінің статистикасын анықтауда профессор А.Қ.Жұбановтың «Задача получения на ЭВМ частотных списков лингвистических единиц» [1] атты мақаласында келтірілген алгоритмдік сызбасы компьютерлік программа жазуға негіз болды. Сонымен бірге, аталған автордың «К вопросу о графемной статистике казахского текста» [2] атты еңбегіндегі әріптердің статистикасы жайлы мәліметтер де осы мақаланы жазуға ықпал етті.

Әдебиет:

1. Жубанов А.К. Задача получения на ЭВМ частотных списков лингвистических единиц. В кн.: Статистика казахского текста. Алма-Ата, 1973. С. 263-299.

2. Жубанов А.К. К вопросу о графемной статистике казахского текста. В кн.: Вопросы казахской фонетики и фонологии. Алма-Ата: Наука, 1979. С. 49-52.

3. Исенгельдина А.А. Факторы, определяющие относительную частотность фонем. В кн.: Статистика казахского текста. Алма-Ата, 1973. С. 659-662.

4. Пиотровский Р.Г. Моделирование фонологических систем и методы их сравнения. М.Л.: Наука, 1966. 299 с.

Т.В.К.К.