

МРНТИ 16.31.02

А.Қ.Жұбанов¹, А.Ә.Жаңабекова¹, Д.Токмырзаев¹

¹А.Байтұрсынұлы атындағы Тіл білімі институтының
бас ғылыми қызметкері, филология ғылымдарының докторы, профессор,
Алматы қаласы, Қазақстан

²А.Байтұрсынұлы атындағы Тіл білімі институтының
бас ғылыми қызметкері, филология ғылымдарының докторы,
Алматы қаласы, Қазақстан

³А.Байтұрсынұлы атындағы Тіл білімі институтының ғылыми
қызметкері, бағдарламашы, Алматы қаласы, Қазақстан

ҰЛТТЫҚ ҚАЗАҚ ӘЛПБИИ ПЕРНЕТАҚТАДА ОРНАЛАСТЫРУДЫҢ СТАТИСТИКАЛЫҚ НЕГІЗДЕРІ

Аннотация: А.Байтұрсынұлы атындағы Тіл білімі институты Қолданбалы лингвистика бөлімі пернетақтада ең жиі қолданылатын әріптерді анықтау үшін қазақ тіліне жиілік талдау жүргізді.

Талдау нәтижесінде қазақ тіліндегі ең жиі кездесетін әріптер мен жиі кездесетін әріп тіркесімдері анықталып, ұлттық пернетақтада әріптерді тиімді орналастыруға қажетті статистикалық ақпарат берілді.

Сонымен қатар осы жиілік талдау нәтижесі бойынша қазақ әріптері мен әріп тіркестерінің жиі кездесуі ескеріліп орналастырылған пернетақтаның пилоттық жобасы ұсынылды.

Мақалада жаңа ұлттық әліпби негізінде латын қаріпті қазақ әріптерін пернетақтада орналастырудың статистикалық негіздері сипатталған. Пернетақтада әріптерді тиімді орналастыру үшін кирилдегі қазақ әріптерінің қолданыстағы кездесу жиілігі, сонымен қатар әріптердің бір-бірімен тіркесіп келудегі жиілігі анықталып, осы жиілік талдау бойынша латын қаріпті қазақ әріптері пернетақтаға орналастырылды. Әріптерді пернетақтада теру ыңғайлылығына қарай пернетақта актив және пассив аймақтарға бөлініп, ортаңғы бөлікке жиі кездесетін әріптер, ал шеткі бөлікке сирек кездесетін әріптер орналастырылды. Мақалада әрбір әріптің статистикалық мәліметі мен пернетақтада орналастырылу себептері нақты баяндалады.

Тірек сөздер: жаңа әліпби, пернетақта, әріп жиілігі, жиілік сөздік, статистикалық талдау

А.К. Жұбанов¹, А.А. Жаңабекова², Д. Токмырзаев³

¹главный научный сотрудник Института языкознания имени А. Байтұрсынова,
доктор филологических наук, профессор,
Алматы, Казахстан

²главный научный сотрудник Института языкознания имени А. Байтұрсынова,
доктор филологических наук, Алматы, Казахстан

³Научный сотрудник Института языкознания им. А. Байтұрсынова,
программист, Алматы, Казахстан

СТАТИСТИЧЕСКИЕ ОСНОВЫ РАСПОЛОЖЕНИЯ НАЦИОНАЛЬНОГО КАЗАХСКОГО АЛФАВИТА НА КЛАВИАТУРЕ

Аннотация. Институт языкознания имени А. Байтұрсынова отдел прикладной лингвистики провел частотный анализ казахского языка для определения наиболее часто используемых букв на клавиатуре.

В результате анализа выявлены наиболее часто встречающиеся буквы на казахском языке и наиболее часто встречающиеся буквенные комбинации, дана статистическая информация, необходимая для эффективного размещения букв на национальной клавиатуре.

Кроме того, по результатам данного частотного анализа был представлен пилотный проект клавиатуры, размещенный с учетом частых встречах казахских букв и буквосочетаний.

В статье описываются статистические основы расположения национального казахского алфавита на клавиатуре. Для удобного расположения букв казахского алфавита на компьютерной клавиатуре были проведены статистические исследования для выявления часто употребляемых казахских букв и их сочетаний как с учетом их места в слове (в начале, в конце слова), а также без их учета. С этой целью сама клавиатура была условно разбита на активный (центральный) участок для высокочастотных букв и на пассивный (крайние) участок для низкочастотных букв казахского языка.

В статье также приводятся статистические данные для отдельных букв казахского языка и даны краткое объяснение причины выбора мест для отдельных букв национального казахского алфавита.

Ключевые слова: новый алфавит, клавиатура, частота букв, частотный словарь, статистический анализ

A.K. Zhubanov¹, A.A. Zhanabekova², D. Tokmyrzaev³

¹Chief researcher of the Institute of Linguistics named after A. Baitursynov,
Doctor of Philology, Professor,
Almaty, Kazakhstan

²Chief researcher of the Institute of Linguistics named after A. Baitursynov,
Doctor of Philology, Almaty, Kazakhstan

³Researcher of the Institute of Linguistics named after A. Baitursynov,
programmer, Almaty, Kazakhstan

THE STATISTICAL FOUNDATIONS OF THE LOCATION OF THE NATIONAL KAZAKH ALPHABET ON THE KEYBOARD

Annotation. The Department of Applied Linguistics of the Institute of Linguistics named after A. Baitursynov conducted a frequency analysis of the Kazakh language to determine the most frequently used letters on the keyboard.

As a result of the analysis, the most common letters in the Kazakh language and the most common letter combinations have been identified, and the statistical information necessary for the effective placement of letters on the national keyboard has been provided.

In addition, based on the results of this frequency analysis, a pilot keyboard project was presented which was developed taking into account the frequency of the use of Kazakh letters and letter combinations.

The article describes the statistical foundations of the location of the national Kazakh alphabet on the keyboard. For the convenient arrangement of the letters of the Kazakh alphabet on a computer keyboard, statistical studies were conducted to identify frequently used Kazakh letters and their combinations, both taking into account their place in the word (at the beginning, at the end of the word), and without taking them into account. For this purpose, the keyboard itself was conditionally divided into an active (central) section for high-frequency letters and a passive (extreme) area for low-frequency letters of the Kazakh language.

The article also provides statistical data for separate Kazakh language letters and gives a brief explanation of the reasons for choosing their places.

Keywords: new alphabet, keyboard, frequency of letters, frequency dictionary, statistical analysis

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. Кирил 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. Ағылшыншадағы бос орындарға қойылған пернетақта

Бұл пернетақтада ағылшындағы бос орындарға жаңаша таңбаланған ә, ң, ғ, у, ү, ө (á, í, ĝ, 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 орналастырылмады. 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. Бұл жиілік талдау Қазақстан Республикасы Білім және ғылым министрлігі тапсырысымен жарық көрген «Жалпы білім берудегі қазақ тілінің жиілік сөздігіне» негіз болған «Қазақ тілінің сөзформалар сөздігінің» мәтіні бойынша жасалды. Бұл сөздікте қазақ тілінің 5 стилінен 7 миллионнан астам сөзқолданыстан тұратын мәтіндер қамтылды. Төменде осы 7 миллионнан астам сөзқолданыстан алынған қазақ әріптерінің жиілік сөздігі берілді.

1-кесте. Қазақ тіліндегі әріптердің жиілікті-әліпбилік 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

1-кестедегі мәліметтер 3 топқа бөліп қарастырылды.

1-топқа ең жиі қолданатын 19 әріп *a, e, ы, н, і, т, р, л, д, с, м, қ, о, к, з, б, й, у, з* енді. Бұл әріптердің ішінде ең жиі қолданыс табатын 1-орынға ие «*a*» әрпі **8630862** рет қолданып, мәтіннің *12,796* пайызын қамтыды, ал жиілік жағынан 2-орынға ие болатын «*e*» әрпі *5660025* рет қолданып, мәтіннің *8.391* пайызын құрады. Осы «*a*», «*e*» әріптерінің біріккендегі қолданысы барлық мәтіннің *21,187* пайызын қамтыды. 1-топқа жататын ең жиі қолданатын 19 әріп (*a, e, ы, н, і, т, р, л, д, с, м, қ, о, к, з, б, й, у, з*) барлық мәтіннің *87,630* пайызын қамтыды.

2-топқа қолдану жиілігі жағынан абсолютті жиіліктері *978676-524073* аралығындағы сан жағынан 10-ға тең (*п, ш, г, ң, и, ж, ұ, ө, ү, ә*) әріптер жатады. Аталған 10 әріп қайталана қолданыла келіп, барлық мәтіннің *11,442* пайызын қамтиды.

3-топқа сирек қолданатын басқа тілдерден енген келесі 13 әріп жатқызылды: *я, х, ц, ф, в, э, ь, ю, һ, 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 қамтиды.

Қазақ тілінің ұлттық компьютерлік пернетақтасында *1-кестедегі* жиі кездесетін 1-топтағы 19 әріп екі бөлікке бөлініп, пернетақтаның ортаңғы бөлігіне орналастырылды. Екіәріптік тіркестердің жиілік сөздігіндегі ең жиі қолданатын, мысалы, мәтіннің 90 пайызын қамтитын рет сандары 1-217 аралығын қамтитын тіркестердің екінші әріптерін *1-кестедегі* рет сандары 1-29 аралығындағы әріптерге тіркесуге пернетақтаның жақын аралығына, яғни сол қол мен оң қол саусақтарына сәйкестендіріп орналастыру мүмкіндіктері қарастырылды.

Ұлттық пернетақта жасауда жеке әріп жиілігі ғана емес, саусақпен оңай әрі тез басу үшін әріптердің тіркесімінің де жиілік сөздіктерін ескеру қажет. Яғни қай әріп қай әріппен жиі тіркесіп қолданылады деген мәселе маңызды. Осы проблеманы шешу үшін қазақ тілінің жоғарыда аталған 7 миллионнан астам сөзқолданыстан тұратын қазақ тілінің мәтіндерінен құрастырылған «Қазақ тілінің сөзформалар жиілік сөздігінен», осы сөздіктің статистикалық деректерін пайдалана отырып, қазақ әріптерінің екіәріптік тіркестерінің жиілік сөздігін құрастырдық.

2-кесте. Қазақ тілінің екіәріптік тіркестерінің жиілік сөздігінен 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		1	2	3
			4				4
1	ар	688045	2,035495	44	ма	229464	46,495744
2	ан	582201	3,757865	45	са	226934	47,167100
3	ал	565089	5,429610	46	ге	226772	47,837976
4	ын	562671	7,094202	47	ел	209952	48,459093
5	да	560939	8,753670	48	ек	208945	49,077231
6	та	511161	10,265877	49	ғы	207327	49,690582
7	ен	501533	11,749600	50	лд	198170	50,276843
8	ер	481450	13,173909	51	ыр	197109	50,859966
9	қа	460416	14,535993	52	ағ	196589	51,441550

10	де	458037	15,891038	53	рд	195039	52,018548
11	ін	451434	17,226550	54	ес	189725	52,579826
12	ла	422562	18,476647	55	іл	189148	53,139397
13	ды	417831	19,712748	56	тг	184683	53,685759
14	ты	404750	20,910150	57	бі	184178	54,230626
15	нд	399330	22,091518	58	ің	181594	54,767850
16	ға	396123	23,263399	59	лі	181226	55,303984
17	ке	371958	24,363790	60	гі	178293	55,831442
18	ай	369359	25,456493	61	ау	177800	56,357441
19	лы	342672	26,470245	62	қы	172081	56,866521
20	ті	333046	27,455520	63	ре	169334	57,367475
21	ді	331481	28,436166	64	ыс	168863	57,867035
22	ас	330440	29,413731	65	сі	168801	58,366412
23	сы	327308	30,382031	66	ып	166290	58,858360
24	ме	323281	31,338418	67	ні	163289	59,341430
25	ат	284740	32,180786	68	ам	161755	59,819962
26	ба	281240	33,012799	69	ік	154050	60,275700
27	ны	276945	33,832107	70	ым	147729	60,712738
28	ың	275833	34,648124	71	ім	147386	61,148761
29	ол	273816	35,458175	72	бо	146149	61,581124
30	те	273269	36,266607	73	қт	144973	62,010009
31	ет	272744	37,073486	74	ад	144878	62,438613
32	ры	263800	37,853906	75	бе	143874	62,864246
33	ық	261904	38,628716	76	шы	140779	63,280723
34	ра	257203	39,389619	77	се	137879	63,688621
35	на	250705	40,131299	78	ем	133338	64,083085
36	жа	248531	40,866547	79	із	131926	64,473371
37	ақ	243641	41,587328	80	кі	130520	64,859499
38	ыл	241848	42,302805	81	ег	128229	65,238848
39	ле	241186	43,016324	82	кө	127963	65,617411
40	не	239816	43,725790	83	ор	126556	65,991811
41	ст	239516	44,434368	84	ап	120653	66,348748
42	ір	237178	45,136030	85	мы	120124	66,704120
43	рі	230151	45,816903	86	ыз	119848	67,058675

Аталған екіәріптік жиілік сөздікте әртүрлі тіркестердің 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. (Бұл сандық мәліметтер әріп тіркесі сөздігінің толық нұсқасында).

Енді осы сөздіктің жоғары жиілікті бөлігінен үзінді ретінде абсолютті жиілігі 100 000-нан жоғары жиілікті зонаға жататын 100 екіәріптік тіркестер бөлігі 2-кестеде берілген.

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 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 2-кестедегі екіәріптік тіркестердің ең жиі қолданылатын тіркестердің екінші әріптерін жиі кездесетін әріптерге тіркесуге пернетақтаның жақын аралығына,

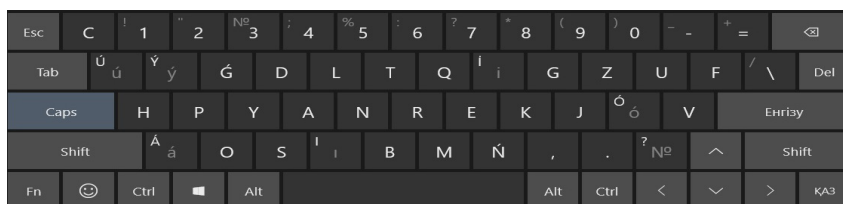
яғни сол қол мен оң қол саусақтарына сәйкестендіріп орналастыру қажет.

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 қолданған әріптер орын алды, ал олардың оң жақтарында екіәріптік тіркестер 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 кирил 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 орналас-тырылған қатардың басуға, яғни мәтін теруге ыңғайлылығына қарай актив және пассив аймақтарға бөлінеді. Пернетақтаның әріп теруге ыңғайлы жері ортаңғы аймақ, әсіресе кертік 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 кертік тұрған ортаңғы қатар, яғни қазіргі кирилдегі Ф әрпінен басталып, Э әрпіне аяталатын қатар. Одан кейінгі актив аймақ осы қатардың үстінде орналасқан үстінгі қатар. Ал астыңғы қатар көбінесе сиректеу кездесетін әріптер орналасатын аймақ.

Батырмалардың тік бағыттағы (вертикалды) межесі кертік тұрған батырмалар-

дың 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 I-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 жатқанымен, тіркесімі бойынша онша 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 (ME). 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 (KE). Тіркесте Е әрпімен жиі кездесетін болғандықтан,

кертінде тұрған Е-нің жанына сол жақ қолға орналастырылды.

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

Б (В) 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.

У (У) әрпі – әріп жиілігінде 18-орында, әріп тіркесінде 61-орында, соңғы (екінші) позицияда (АУ). 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 косәріп (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.

Н (N) әрпі – әріп жиілігі бойынша 23-орында, яғни сирек, алайда әріп тіркесінде 28-орын тым сирек кездесетіндерге жаппайды. Сондықтан осы екі жағдайда да ескеріліп, астыңғы шеткі, алайда жиі кездесетін Е-нің астында 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.

Ұ (U) әрпі – әріп жиілігі бойынша 26-орында, сирек кездесетіндерге жатады. Әріп тіркесінде де 87-орында бастапқы позицияда (ҰР) 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 қатарда орналастырылды.

Ү (Ú) 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/Һ (H) әріптері – өте сирек. Х – әріп жиілігінде 31-орында, Һ – 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 астында 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 сыйымдылығы мен қуат сыйымдылығына қарай сынақтан өткізілуде.

Сонымен, егер жиілік ұстанымын негізге алатын болсақ, 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 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 (сөз басындағы, сөз соңындағы) жағдайына да мән берілуі керек.

ӘДЕБИЕТТЕР ТІЗІМІ:

1. <http://stud24.ru/information/istoriya-sozdaniya-klaviatury/22660-65009-page1.html>
2. <http://encicl.narod.ru/klav.htm>

МРНТИ 16.21.37

З.М.Базарбаева

А.Байтұрсынұлы атындағы Тіл білімі институтының бас ғылыми қызметкері, филология ғылымдарының докторы, профессор
Алматы қаласы, Қазақстан

КІРМЕ СӨЗДЕРДІҢ ҚАЗАҚ ТІЛІНІҢ ФОНОЛОГИЯЛЫҚ ЗАҢДЫЛЫҚТАРЫНА ИКЕМДЕП ЖАЗУДЫҢ ПРИНЦИПТЕРІ

Аннотация: Кірме сөздерді қазақ тілінің фонологиялық заңдылықтарына икемдеп жазу және оның принциптерін жан-жақты қарастыру қазіргі кезде маңызды болып отыр. Кірме әріптермен жазылатын сөздердің фонетика-фонологиялық негіздері талданып беріледі. Қазіргі кезде кейбір кірме сөздерді игеріп, қазақ тіліне икемдеп «сындырып» айту мен жазу оңайға түспейді. Осыған орай соңғы кезде латынға көшумен байланысты кірме сөздерді таңбалауда бірнеше көзқарас бар. Осы орайда кірме сөзді қалай жазу керектігі мақалада мысал арқылы беріледі.

Мақалада кірме сөздердің қазақ тілінің фонологиялық заңдылықтарына икемдеп жазудың принциптері жан-жақты сипатталып, оны берудің жолдары беріледі.

Тірек сөздер: фонология, фонетика, фонема, орфография.