

## КОНВЕРТОР-АКУТ

Әліпби ауыстыруда ең алдымен, осындай сөздердің емлесін нақтылап алуымыз қажет. Бұл әсіресе, кирилл қаріпті көптеген әдебиеттерді (мәтіндерді) автоматты жолмен латынға аудару үшін басты шешіп алатын мәселе.

Кирилл қарпіндегі мәтіндерді латын қарпіне аудару оп-оңай жүзеге асырылмайды. Өйткені әліпби ауыстыру бір әріптің орнына бір әріпті қоя салу емес, бұл лингвистикалық қана емес, саяси-әлеуметтік те реформа.

Жазу реформасын жүзеге асыруда әртүрлі әлеуметтік көзқарас туғызып жүрген күрделі мәселелердің бірі – кирилл қарпінде жазылған жазба әдебиеттерімізді жоғалтып аламыз ба деген қауіп. Қазіргі компьютерлік техниканың дамыған заманында бұл мәселенің шешімін табу көп қиындыққа түспейтіні белгілі. Арине, кітап түрінде қайта шығару сияқты шығындары болуы мүмкін, ал электронды нұсқада әдебиеттерді, жазба мұраларымызды сақтап қалуға мүмкіндік әбден бар. Мұны автоматты аударым жасау ісі деп атайды. Мұндай амал-әрекетті жүзеге асырушы бағдарламаны «конвертер» деп атайды. Кейде орысша «перекодировчик» те қолданады.

Кирилл қарпіндегі мәтіндерді латын қарпіне автоматты аударым жасайтын конвертерлер әріптің орнына әріпті ауыстыра салмайды. Егер олай етсек, қаншама көлемді әдебиеттеріміз емле ережелері ескерілмеген қате орфограммаланады. Қазіргі интернет сайттарында жүрген конвертер бағдарламаларының көпшілігі осылай емле ережелері ескерілмей жасалған деуге болады. Автоматты аударым жасауға арналған бағдарлама жасау ол тек программистің жұмысы емес, ол емле ережелерін жіті қадағалап отырған лингвистердің де жұмысы екендігін айту қажет. Лингвистердің қадағалуынсыз автоматты аударым жасайтын компьютерлік бағдарлама мәтіндерді дұрыс көшірім жасай алмайды. Бұл орайда конвертер жасаушы бағдарламашыға тілшілер тарыпынан мынадай лингвистикалық нұсқаулық (программаға арналған емле) дайындалады.

**Бірінші қадам.** Компьютерлік бағдарлама үшін ең алдымен кирилл және соған сәйкес латынша таңбалары көрсетілген *кесте 1* беріледі.

Кесте 1 – Кирилл және латын қаріптері сәйкес берілген әліпби

Латын графикасына негізделген қазақ тілі  
ӘЛПБИ

№	Жазылуы	Дыбысталуы	№	Жазылуы	Дыбысталуы
1	A a	[a]	17	Ń ń	[ң]
2	Á á	[ə]	18	O o	[o]
3	B b	[b]	19	Ó ó	[ө]
4	D d	[d]	20	P p	[п]
5	E e	[e]	21	Q q	[қ]
6	F f	[ф]	22	R r	[р]
7	G g	[г]	23	S s	[с]
8	Ĝ ĝ	[ғ]	24	T t	[т]
9	H h	[x], [h]	25	U u	[ұ]
10	I i	[i]	26	Ú ú	[ү]
11	I ı	[и], [й]	27	V v	[в]
12	J j	[ж]	28	Y y	[ы]
13	K k	[к]	29	Ý ý	[у]
14	L l	[л]	30	Z z	[з]
15	M m	[м]	31	Sh sh	[ш]
16	N n	[н]	32	Ch ch	[ч]

Жоғарыдағы суретте қазақ тіліндегі 32 әріп берілген, **h** және **x** әріптері бір таңбамен таңбаланады – **Hh**. Сол сияқты әліпбиде кирилдегі **И** және **Й** әріптері латында бір ғана **I** әрпімен таңбаланған. Кестеде кирилдегі жалпы саны – 34 әріп тұр. Кирилдегі **Ёё, Цц, Щщ, Ъ, ъ, Ээ, Юю, Яя** әріптері таңбаланбаған. Жоғарыдағы 34 әріпке осы 8 әріпті қоссақ, кирилдегі 42 әріп болады.

**Ш** және **ч** әріптері диграфтармен берілген – **Ш- Sh; Ч- Ch**.

Мәтінде автоматты түрде кирилден латынға ауыстыру үшін кирилдегі барлық 42 әріптің қалай таңбаланатыны туралы нұсқаулық (емле) берілу керек. Сол үшін әліпбиде таңбаланбаған әріптерді латында қалай таңбаланатыны туралы да программаға арналған емле жасалады. Программаға арналған емле дайындау орфографиялық емле ережелеріне сүйенеді. Біздің конвертор қазіргі орфографиялық топтың жасаған емле ережелерінің соңғы үлгісін басшылыққа алады [1].

**Екінші қадам.** Жоғарыда берілген латынға негізделген ұлттық әліпбиден басқа бұл әліпбиге енбеген 8 әріптің қалай таңбаланаты туралы нұсқаулық *кесте 2* дайындалады.

Кесте 2 – Ұлттық әліпбиге енбеген 8 әріптің орфограммалануы

1	Ёё		Ee
2	Цц		S
		<i>пицца, ницца</i> деген екі сөзде ғана	ts
3	Щщ	кірме сөздерде	sh
		ал қазақтың байырғы сөздерінде, <i>ащы, тұщы</i>	shsh
4	Ь		-
5	ъ		-
6	Ээ		Ee
7	Юю		iy
8	Яя		ia

Кестеден көрініп тұрғандай, әліпбиде жоқ **Ь, ъ** әріптері латын қаріпті қазақ мәтіндерінде таңбаланбайтын болады. Программа жіңішкелік және айыру (жуан) белгілерін жоғарыдағы нұсқаулық кестеге сәйкес автоматты түрде алып тастап отырады.

**Э** және **Ё** әрпі бар сөздер түгелімен **E** таңбасына ауыстырылады. **Ю** әрпі – **iy** болып, сол сияқты **Я** әрпі – **ia** болып екі әріппен таңбаланатындығы орфографиялық емле ережелерінде бекітіліп отыр. Осы ретте программа да осы нұсқаулық кестеге сәйкес мәтінде кездескен **Я, Ю** әріптерін екі таңбамен аударым жасайтын болады.

Қазақ тілінде **Ц** әрпі тек қана кірме сөздерде, терминдерде ғана кездеседі. Бұл әріп ұлттық әліпбиге енгізілмеген. Алайда бұл әріп терминдерде кездесетіндіктен, орфографиялық емле жасаушылар оны қалайда таңбалау мақсатында ұлттық әліпбидегі **S** әрпімен беруді жөн санап отыр. Қазақ тілінде термин ретінде қолданысқа еніп кеткен *пицца, ницца* сөздерінде **Ц** әрпі екі рет кездесетіндіктен, мұндай жағдайда **ts** диграфымен беру көрсетіледі. Ал қалған жағдайлардың барлығында **Ц** әрпі бір ғана **S**-мен таңбаланады. Осы орфографиялық емлеген сәйкес, жоғарыдағы кестедегі нұсқаулықта көрсетілгенде, конвертер бағдарламасы *пицца* және *ницца* сөздерінен басқа мәтінде кездесетін **Ц** әрпі бар сөздерде оны **S**-ға көшіреді. Программаға *пицца* және *ницца* сөздері қалыс жағдай (исключение) ретінде көрсетіледі.

**Щ** әрпі қазақ тілінің байырғы сөздерінде тек *ащы, тұщы* сөздерінде ғана қолданылады. Ал негізінен орыс тілінен енген кірме сөздерде кездеседі. Емле жасаушы тілші-мамандар *ащы, тұщы* сөздерінде **Щ**-ны екі қосарлама диграфпен (**shsh**) беруді, ал кірме сөздерде бір ғана диграфпен (**sh**) беруді шешіп отыр.

Бұл жоғарыда айтылғандардың барлығы негізінен әліпбидегі әріптерді таңбалауға қажетті нұсқаулықтар болып табылады. Яғни кестеге салынған таңбалар арқылы программа кирил қаріптерін латын графикасына ауыстырып шығады.

**Үшінші қадам.** Мәтінде автоматты аударым жасауда жоғарыда айтылғандан басқа да программаға арналған емле ережелерін жасау қажет. Солардың бірі – таратып жазу шешіліп отырған **Я** және **Ю** әріптеріне қатысты.

Қазақ тілінде **я, ю (йа, йу (йүу))** төл әріптері болмаса да, кезінде кирилл қарпінің енуімен бірге, бір әріппен таңбаланған болатын. Қазір латын графикасына көшіруде бұл әріптер екіге ажыратылып таңбаланатын болғандықтан, осы екі әріп кездесетін кейбір сөздерді автоматты түрде латынға ауыстыруда мынадай жағдайларды ескеру керек болады. Мәселен, бұл екі әріптің құрамында да **и** әрпі бар және қазақ тілінде осы **я, ю** әріптерінің алдынан **и** әрпі жазылатын сөздер бар. Егер **я, ю** әріптерін **иа – ia, иу – iû** деп таратып жазатын болса, мұндай сөздерде екі **и (i)** әрпі қатар келіп қалады. Мұндай жағдайда орфографист-мамандар емледе бір **и**-ді ғана қалдыруды ұсынып отыр. Осыған сәйкес программаға мынадай ереже беріледі:

*«Егер кирилл қарпіндегі мәтіндерде **я** әрпі бар сөздердің алдында **и** әрпі кездессе, латын қарпіне ауыстырғанда бір ғана **и (i)** таңбаланады».*

Мысалы: **қиял**. Егер біз осы сөзді автоматты түрде латынға аударым жасайтын болсақ, **qпal** болып орфограммаланған болар еді. Мұндай сөздерде **и (i)** әрпін екі рет жазудың қажеті болмайды. Бұл сөздің дұрыс латынша таңбалану нұсқасы **qial** болады. Осы сияқты **ия** әріптерімен келген сөздердің барлығын конвертор **ia** етіп таңбалайтын болады.

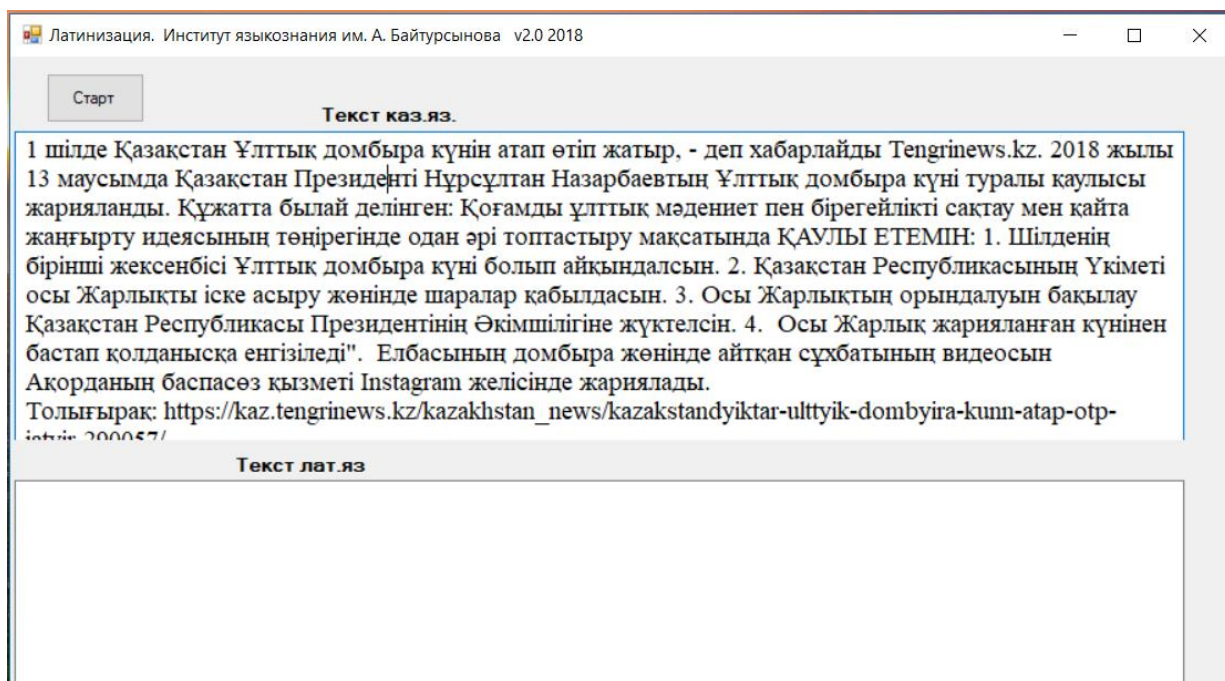
Осы сияқты **ю** әрпіне қатысты мынадай ереже-нұсқаулық беріледі:

*«Егер кирилл қарпіндегі мәтіндерде **ю** әрпі бар сөздердің алдында **и** әрпі кездессе, латын қарпіне ауыстырғанда бір ғана **и (i)** таңбаланады».*

Мысалы: **қию**. Егер біз осы сөзді автоматты түрде латынға аударым жасайтын болсақ, **qпiû** болып орфограммаланған болар еді. Мұндай сөздерде **и (i)** әрпін екі рет жазудың қажеті болмайды. Бұл сөздің дұрыс латынша таңбалану нұсқасы **qiû** болады. Осы сияқты **ию** әріптерімен келген сөздердің барлығын конвертор **iû** етіп таңбалайтын болады.

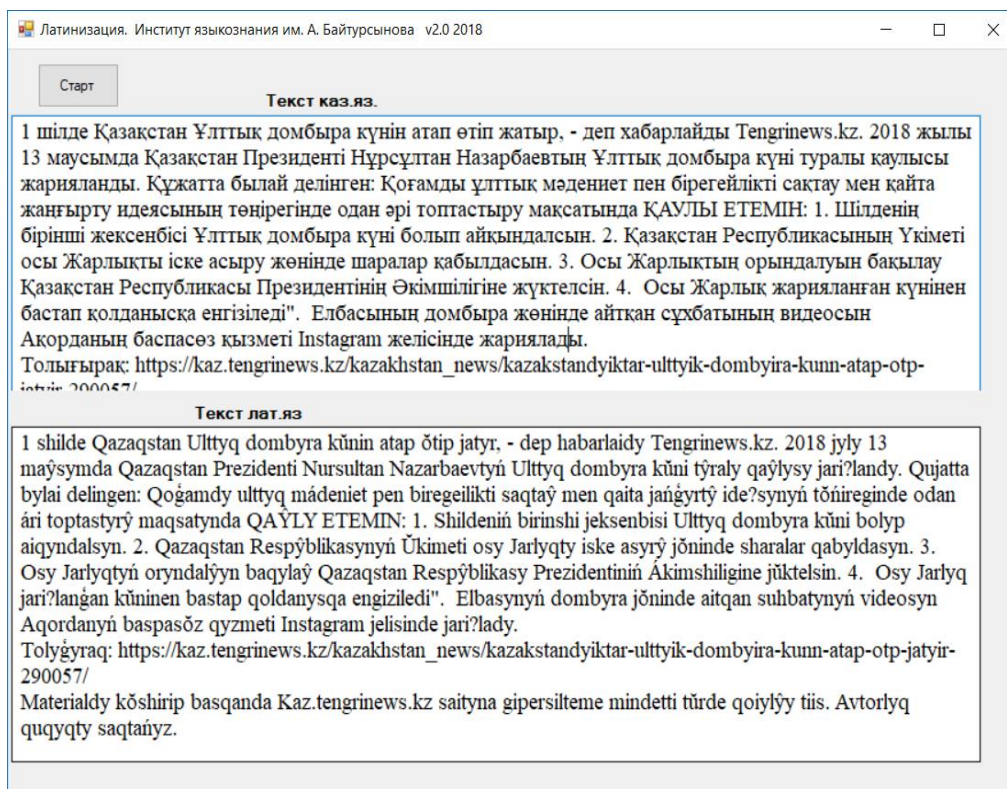
Программа әзірлемесі Visual Studio 2017 ортасында, программалаудың C# тілінде жасалды.

Программаны енгізер кезде екі терезесі бар мәтін енгізу экраны ашылады. *Сурет 1.*



Сурет 1 – Кирилл қарпінен автоматты аударым жасайтын компьютерлік бағдарлама терезесі

Жоғарғы терезеге латынға көшірілуге тиісті кирилл қаріпті қазақ мәтіні енгізіледі. Содан соң «СТАРТ» түймешесі басылады. Осы кезде астыңғы терезеден латын қ аріпті қазақ мәтіні шығады. *Сурет 2.*



Сурет 2 – Латынға көшіру программасы

Бұл бағдарлама v1.0 программасының бірінші нұсқасы болып табылады. Келешекте қазақ тіліндегі емле ережелеріне сәйкес жетілдіріледі, сонымен қатар осы бағдарлама бойынша көлемді файлдарды салу мен көшірудің бағдарламалары жасалады.

Сонымен тілтанушы мен бағдарламашының бірлескен жұмысының нәтижесінде соңғы акут әліпбиімен соңғы (қазіргі) жасалып жатқан емле-ережелері негізге алынған кирилл қарпіндегі қазақ мәтіндерді автоматты түрде латын графикасына ауыстыратын ІТ-қосымша жасалды. Бағдарлама институттың латын бойынша жоспарларын жүзеге асыруға өз септігін тигізуде. Программамен (конвертор) төмендегі сілтеме бойынша танысуға болады.

<https://e.mail.ru/thread/0:15388085190000000774:0/>

Қорыта келе, программаға берілетін емле ережелерін қысқаша былай көрсетуге болады:

1) Әліпбидегі сайма-сай келетін қазақ әріптерінің кирилл және латын нұсқасы;

2) Алынып тасталатын әріптер, мысалы, *ь, ъ*;

3) Диграфтармен жазылатын, яғни қос әріптермен таңбаланатын әріптер;

4) Кейбір әріптердің сөз басында, сөз ортасында позициялық тұрғыдан екі түрлі таңбалануы;

5) Кирилдегі бір әріппен таңбаланатын кейбір дыбыстардың қос әріппен таратылып жазылуы, мысалы: *ю – иу; я – иа*;

6) Орфографиялық емле ережелеріне сәйкес кейбір қос әріпті сөздерде қайталанатын дауыссыз дыбыстың түсіп қалуы, мысалы: *метал*;

7) *Ю, я* сияқты таратылып жазылатын дыбыстармен келген сөздерде осы әріптер алдында и әрпі келген жағдайда бір и таңбасын түсіріп қолдану және автоматты аударымда осыны ескеру т.б.

Конвертор жұмысын осымен толық бітті дей алмаймыз. Себебі конвертор алдағы уақытта әліпби, емледегі өзгерістерді ескеріп, қайта жаңартылып отырады. Сонымен қатар мәтіндерді автоматты түрде аударым жасауда проблемалар шығуы мүмкін. Алайда қазіргі біздің жасаған кирилл қаріпті мәтіндерді латын қарпіне автоматты түрде аударым жсайтын конвертор соңғы емле ережелеріне сүйеніп жасалып отыр.