

дыбыстардың, өзге маркерлердің екінші жергілікті тіл жазуында еш мәні жоқ. Сондықтан романдаудың «транслитерациялық кілті» қажет. Ал романдау кілтін түсіндіру үшін транскрипциялық әдіске жүгінуге тура келеді, өйткені түсінік-теме өзге тіл иелменіне өз тілінде транскрипция арқылы берілуі тиіс. Яғни транслитерация мен транскрипция өзара тығыз байланысты тәсілдер.

Сондықтан ономастикалық атауды романдаудың үш тәсілі: транслитерация, транскрипция және аударма тілші маманнан жан-жақты біліктілікті қажет етеді.

ӘДЕБИЕТТЕР ТІЗІМІ:

[1] Жаңа әліпби негізіндегі қазақ тілі емлесінің ережелері // Егемен Қазақстан, 18 қаңтар 2019 жыл

МРНТИ 16.31.21

А.Ә.Жаңабекова¹, Қ.Қойбағаров²

¹А.Байтұрсынұлы атындағы Тіл білімі институты
Қолданбалы лингвистика бөлімінің меңгерушісі, филология
ғылымдарының докторы, Алматы қаласы, Қазақстан

²Қолданбалы лингвистика бөлімінің программисі,
Алматы қаласы, Қазақстан

КИРИЛЛ ҚАРПІНДЕГІ МӘТІНДЕРДІ ЛАТЫН ГРАФИКАСЫНА КӨШРУДЕГІ ЕМЛЕ ЕРЕЖЕЛЕРІ ЖӘНЕ ПРОГРАММАЛЫҚ ҚАМТАМАСЫЗ ЕТУ ӘЗІРЛЕМЕСІ

Аннотация: 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. 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. Мәтіндерді автоматты аударма жасау үшін, сөздердің емлесі программа үшін аса қажет. Мақала кирилл қарпінен латын графикасына аударым жасатын компьютерлік бағдарлама жұмысына арналған.

Қолданбалы лингвистика бөлімі әліпби бекіту процесі кезеңінде әр жылдары әліпби нұсқаларына сәйкес конвертерларды жасап отырды. 2017 жылы апострофпен жасалған компьютерлік бағдарламаны, 2018-де соңғы ақутқа негізделген жаңа әліпбимен қайта жасады. Бұл орайда конвертер жасаушы бағдарламашыға тілшілер тарыпынан лингвистикалық нұсқаулық (программаға арналған емле) дайындалады. Мақалада осы лингвистикалық нұсқаулық пен конвертор жұмысына сипаттама беріледі.

Тірек сөздер: қаріп ауыстыру, емле, конвертор, автоматты аударым, программаға арналған нұсқаулық

А.А. Жанабекова¹, К. Койбағаров²

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

²Института языкознания имени А. Байтұрсынова, программист отдела
Прикладная лингвистика, Алматы, Казахстан

ПРАВИЛА ПЕРЕВОДА НА ЛАТИНСКУЮ ГРАФИКУ ТЕКСТОВ НА КИРИЛЛИЦЕ И РАЗРАБОТКА ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Аннотация. Одна из сложнейших проблем, вызывающих различные социальные взгляды в осуществлении реформы письменности, – угроза того, можем ли мы потерять наши Писания, написанные в черте Кирилла. Как известно, в эпоху развития современной компьютерной техники решение этой проблемы не может быть затруднено. В электронном варианте есть возможность сохранить литературу, записное наследие. Это называется автоматическим переводом. Программа, реализующая такое действие, называется “конвертером”. Для автоматического перевода текстов, написание слов очень важно для программы. Статья посвящена работе с компьютерными программами, переводящими из шрифта Кирилла на латинскую графику.

Отдел прикладной лингвистики в процессе утверждения алфавита разработали нескольких версии компьютерной программы автоматического перевода текстов кириллицы на латинскую графику. В 2017 году разработали программу с апострофами, а в 2018 году с акутами. С целью разработки конвертора лингвисты подготовили инструкцию для программы. В статье описывается структура инструкций и работаконвертора.

Ключевые слова: переход графики, правила, конвертор, автоматический перевод, инструкция для программы

А.А. Zhanabekova¹, K. Koibagarov²

¹Chief Researcher of the A. Baitursynov Institute of Linguistics,
Doctor of Philology, Almaty, Kazakhstan

²A. Baitursynov Institute of Linguistics,
programmer of the department of applied linguistics,
Almaty, Kazakhstan

RULES FOR THE TRANSLATION OF CYRILLIC TEXTS INTO LATIN GRAPHICS AND SOFTWARE DEVELOPMENT

Annotation. One of the most difficult problems causing different social attitudes in

the implementation of the reform of writing is the threat that we could lose our Scriptures, written within the Cyril line. As you know, in the era of development of modern computer technology, solving this problem cannot be difficult. In the electronic version there is an opportunity to preserve literature, written heritage. This is called automatic translation. The program that implements such an action is called a “converter”. For automatic translation of texts, the spelling of words is very important for the program. The article is devoted to working with computer programs that translate from Cyril’s font to Latin graphics.

The department of applied linguistics in the process of approval of the alphabet developed several versions of a computer program for automatically translating Cyrillic texts into Latin graphics. In 2017, we developed a program with apostrophes, and in 2018 with acute. In order to develop a converter, linguists have prepared instructions for the program. The article describes the structure of instructions and the operation of the converter.

Keywords: transition graphics, rules, converter, automatic translation, instructions for the program

Мәтіндерді автоматты аударма жасау үшін, сөздердің емлесі программа үшін аса қажет. Түптеп келгенде, әліпби ауыстыруда ең алдымен, осындай сөздердің емлесін нақтылап алуымыз қажет. Бұл әсіресе, кирил қаріпті көптеген әдебиеттерді (мәтіндерді) автоматты жолмен латынға аудару үшін басты шешіп алатын мәселе. Әрі келешек жазу емлесін ұлттық дыбыстау жүйесіне бейімдеу өзекті 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. Әсіресе, әліпби ауыстыру кезінде әр ұлттың бітім-болмысына, табиғатына сәйкес фонды беру қиынның қиыны болып табылады. Қазіргі 70 жылдан астам тарихы бар Қазақстандағы кирилл жазуын өзгерту, қоғамдық санаға сіңген таңбалық стереотиптерді өзгерту оңай емес. Кейбір таңбалар бұрыннан ағылшын пернетақтасында пайдаланып жүрген үйреншікті болып, көзшалымға жатық, қабылдауға жеңіл болса, кейбір таңбалар кездейсоқтық туғызып, қабылдауға, жаттауға қиындық тудырады. Сонымен қатар әліпбидегі қазақ тілінде бұрын-соңды кирилл қарпінде болмаған қос әріппен таңбалау тұстары да жазудың эстетикалық айшығына, орфограммалық тұрпатына, яғни көзшалымға сәтсіздік тудырады. Әліпбиден бөлек, жазуда тағы бір атқарылатын үлкен жұмыс – емле ережелерін жасау. Жаңа емле бұрынғы емлені латын қарпіне ауыстыра салу емес, ол бұрынғы емледегі кейбір кемшіліктер мен олқылықтардың орнын толтыру үшін және де кирилл қарпінің экспансиясы ықпалымен ұлттық кодтан айрылып бара жатқан орфограммаларды қайта қалпына келтіру үшін немесе ұлттық сөйлеу жүйесін нормалау үшін де қайта жаңғыртылып жасалуы қажет.

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. Әрине, 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. Осы мәселені сәтті шешу үшін Қолданбалы лингвистика бөлімі әліпби бекіту процесі кезеңінде әр жылдары әліпби нұсқаларына сәйкес конвертерларды жасап отыруды міндет етіп қойып отыр. 2017 жылы апострофпен жасалған компьютерлік бағдарламаны жасаған болатын. 2018-де соңғы ақутқа негізделген жаңа әліпбимен қайта жасады. Бұл орайда конвертер жасаушы бағдарламашыға тілшілер тартыпанын мынадай лингвистикалық нұсқаулық (программаға арналған емле) дайындалады.

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

Латын графикасына негізделген қазақ тілі
ӘЛІПБИ

№	Жазылуы	Дыбысталуы	№	Жазылуы	Дыбысталуы
1	A a	[a]	17	Ń ń	[ɲ]
2	Á á	[ə]	18	O o	[o]
3	B b	[b]	19	Ó ó	[o]
4	D d	[d]	20	P p	[p]
5	E e	[e]	21	Q q	[k]
6	F f	[f]	22	R r	[r]
7	G g	[g]	23	S s	[s]
8	Ĝ ĝ	[ɣ]	24	T t	[t]
9	H h	[x], [h]	25	U u	[ʏ]
10	I i	[i]	26	Ú ú	[y]
11	Ĭ ĭ	[ɨ], [ï]	27	V v	[v]
12	J j	[ʒ]	28	Y y	[ɣ]
13	K k	[k]	29	Ý ý	[y]
14	L l	[l]	30	Z z	[z]
15	M m	[m]	31	Sh sh	[ʃ]
16	N n	[n]	32	Ch ch	[tʃ]

Жоғарыдағы суретте қазақ тіліндегі 32 әріп берілген, *h* және *x* әріптері бір таңбамен таңбаланады – *Hh*. Сол сияқты әліпбидекирилдегі *И* және *Й* әріптері

латында бір ғана *I* әрпімен таңбаланған. Кестеде кирилдегі жалпы саны – 34 әріп түр. Кирилдегі *Ёё, Цц, Щщ, ь, 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- Ч*.

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 [1]. Біздің конвертор қазіргі орфографиялық топтың жасаған емле ережелерінің соңғы үлгісін басшылыққа алады.

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

2-кесте. 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 – *y* 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.

Емле ережелерін жасау кезеңінде мамандар *Ц* әрпін сөз басында бір *S*-мен, сөз ортасында диграф (*ts*) етіп (қос әріппен) таңбалауды да жоспарлаған болатын. Бірақ бұл кейін өзгертілгендіктен, біздің программаға да соған сәйкес өзгеріс енгізілді.

Щ әрпі қазақ тілінің байырғы сөздерінде тек *ащы*, *тұщы* сөздерінде ғана қолданылады. Ал негізінен орыс тілінен енген кірме сөздерде кездеседі. Емле жасаушы тілші-мамандар *ашы*, *тұшы* 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 кирилл 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* (*йұ*)) 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* – *иа*, *иу* – *иу* 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, *qial* 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, *qiu* болып орфограммаланған болар еді. Мұндай сөздерде *и* (*i*) әрпін екі рет жазудың қажеті болмайды. Бұл сөздің дұрыс латынша таңбалану нұсқасы *qiu* болады. Осы сияқты *ию* әріптерімен келген сөздердің барлығын конвертор *iу* етіп таңбалайтын болады.

Программа әзірлемесі Visual Studio 2017 ортасында, программалаудың C# тілінде жасалды. Бұл программа латынға көшірудің 2-ші мәрте жасалған нұсқасы болып табылады. Кирилл қаріптерінен латын қарпіне автоматты аударым жасауда DIC сөздігі жасалды. Мұнда кирилл қарпіндегі әрбір әріптің латын баламасы екінші кестеде сәйкестендіріліп беріледі. *Кесте 3.*

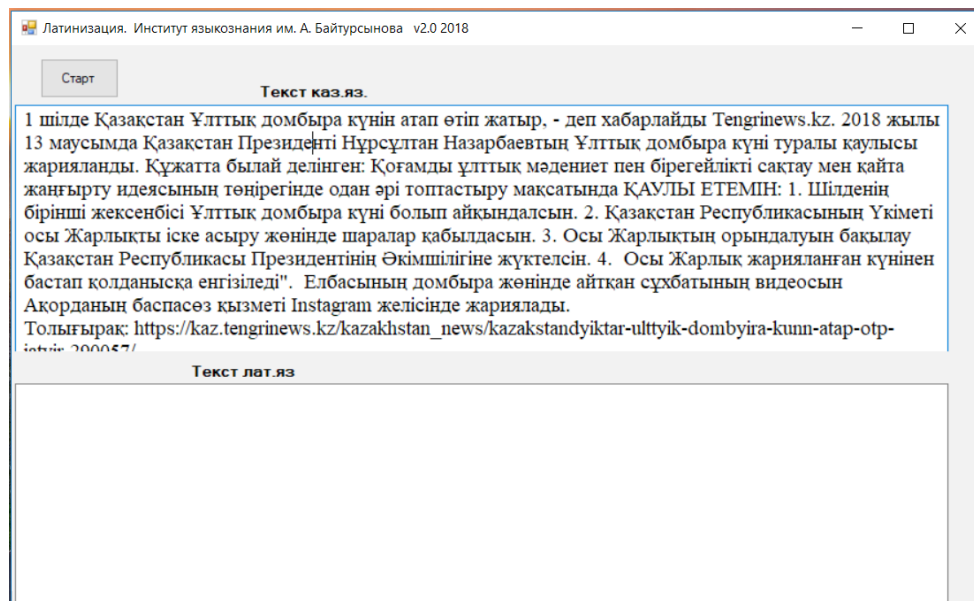
3-кесте. Қазақ және латын қаріптерінің сәйкестендірілген сөздігі

dic.Add("а", "a");	
dic.Add("б", "b");	dic.Add("э", "e");
dic.Add("в", "v");	dic.Add("ю", "iy");
dic.Add("г", "g");	dic.Add("я", "ia");
dic.Add("д", "d");	dic.Add("ы", "y");
dic.Add("е", "e");	dic.Add("й", "i");
dic.Add("ё", "e");	dic.Add("ү", "y");

dic.Add("ж", "j");	dic.Add("ь", "y");
dic.Add("з", "z");	dic.Add("ъ", "x");
dic.Add("и", "i");	dic.Add("«ә», «a»);
dic.Add("к", "k");	dic.Add("і", "i");
dic.Add("л", "l");	dic.Add("«ң», «n»);
dic.Add("н", "n");	dic.Add("«ғ», «g»);
dic.Add("м", "m");	dic.Add("«қ», «q»);
dic.Add("о", "o");	dic.Add("ұ", "u");
dic.Add("п", "p");	dic.Add("к", "g");
dic.Add("р", "r");	dic.Add("ө", "o");
dic.Add("с", "s");	dic.Add("һ", "h");
dic.Add("т", "t");	dic.Add("х", "h");
dic.Add("ф", "f");	dic.Add("ц", "c");
dic.Add("ч", "ch");	dic.Add("ш", "sh");

Программа алгоритмі төмендегіше жұмыс жасайды. Енгізілген мәтін алдымен сөйлемдерге бөлінеді, содан кейін сөйлемдер сөздерге, ең соңынан сөздер әріптерге бөлінеді. Осыдан кейін әрбір әріп латындағы сәйкес қаріппен ауыстырылады. Программа алгоритмі келешекте ауыстыру ережелері бойынша жетілдіріледі.

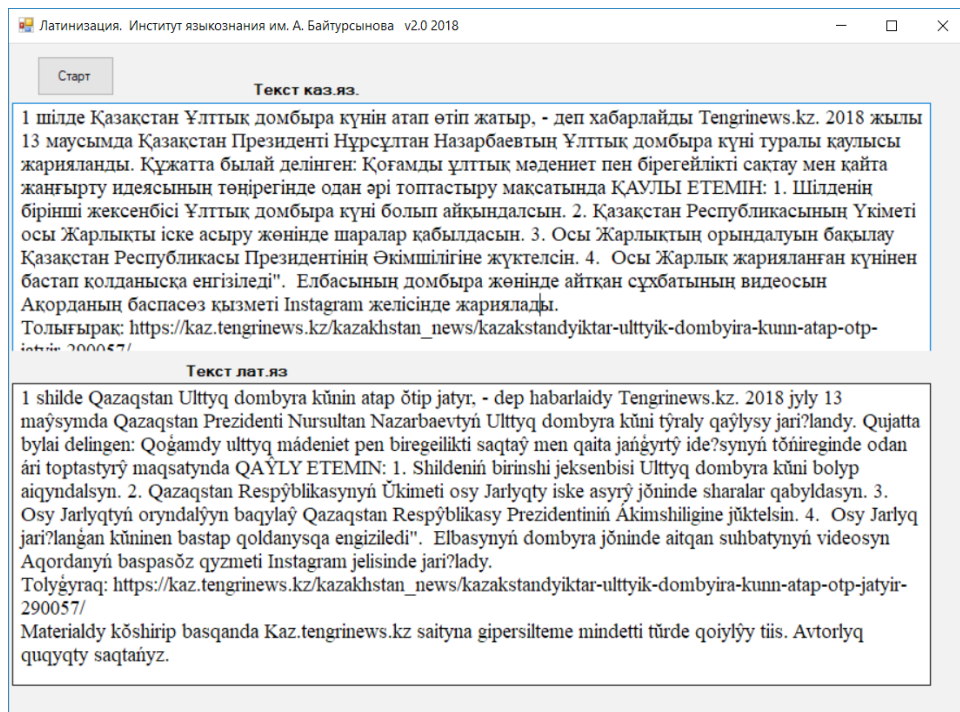
Программаны енгізер кезде екі терезесі бар мәтін енгізу экраны ашылады. Сурет-1.



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 қаріпті қазақ мәтіні шығады. 2-сурет.

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.



Сурет 2 – Латынға көшіру программасы

Сонымен тілтанушы мен бағдарламашының бірлескен жұмысының нәтижесінде соңғы ақут 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.

ӘДЕБИЕТТЕР ТІЗІМІ:

[1]. Жаңа әліпби негізіндегі қазақ тілі емлесінің ережелері. – Астана, 2018.

МРНТИ 16.21.27

Ә.Қоспағарова¹, Л.Аманқұлова²

^{1,2}Қазақ мемлекеттік қыздар педагогикалық университетінің магистрі, аға оқытушылары, Алматы, Қазақстан

БҰҚАРАЛЫҚ МЕРЕКЕЛЕРДІҢ ТОЙЛАНУЫ МЕН САХНАЛАНУЫ

Аннотация: Бұл мақалада бұқаралық мерекелердің сахналау түрлері мен ұйымдастыру бағыттары қарастырылып, болашақ ұйымдастырушы педагогтың