

УДК 004.8

У. А. ТУКЕЕВ, М. А. БОЛАТБЕК, М. А. ИЛЖАНОВ, А. Х. РАЗАХОВ

(<sup>1</sup>әл-Фараби атындағы Қазақ ұлттық университеті,  
«Математика және механика» ғылыми зерттеу институты, Алматы, Қазақстан)

## ҚАЗАҚ-ОРЫС МАШИНАЛЫҚ АУДАРМАСЫ КӨПМӘНДІ БЕЙНЕЛЕУ АППАРАТЫ НЕГІЗІНДЕ

**Аннотация.** Зерттеу жұмысының мақсаты қазақша-орысша машиналық аудармасын көпмәнді бейнелеу аппараты негізінде құрудың үлгілерін, алгоритмдері мен бағдарламасын құрастыру және Қазақстан халқы мәтінді сапалы түрде аударатын машиналық аударма бағдарламасымен қамтамасыз ету болып табылады. Жұмыс нәтижесі ретінде қазақ тіліндегі сөздерді, сөз тіркестері мен қарапайым сөйлемді орыс тіліне сапалы түрде аударатын машиналық аударма бағдарламасы құрастырылды. Жұмыс нәтижесін – мақала тақырыбына сай құрастырылған машиналық аударма жасау бағдарламасын аудармашы көмегіне жүгінуге мәжбүр болған әр адам пайдалана алады. Зерттеудің ғылыми жаңалықтары – екі тілдің грамматикалық сәйкестіктерін орнату кесте түрінде жинақталған қосымшаларға сұраныс жасау арқылы шақыртуға негізделді, енгізілген сөздің түбірі мен қосымшасын ажыратуда іздеу екі жақтан жүргізілді, деректер базасы 9 сөз тобына бөлінген 12 000-нан астам сөзден тұрады.

**Тірек сөздер:** машиналық аударма, машиналық аударма анализаторы, машиналық аударма генераторы, көпмәнді бейнелеу аппараты.

**Ключевые слова:** машинный перевод, анализатор машинного перевода, генератор машинного перевода, аппарат многозначных отображений.

**Keywords:** machine translation, analyser of machine translation, generator of machine translation, set-valued mapping.

**Кіріспе.** Қазіргі таңда отандық нарықта қазақша-орысша және кері бағытта жұмыс істейтін бірнеше машиналық аударма түрлері бар. Алайда олардың көпшілігі сапалы аударма жасай бермейді. Сондықтан да машиналық аударма саласында машиналық аударманың сапасы өзекті мәселе болып қалуда.

Бұл мақалада бір тілден екінші тілге машиналық аударманың тиімді технологиясы ретінде көпмәнді бейнелеу аппаратын пайдалану ұсынылады [1]. Көпмәнді бейнелеу аппараты соңғы 30 жыл ішінде, атап айтсақ, ойындар теориясында, экстремальды есептер теориясында, математикалық экономикада қарқынды дамып келеді [2].

**Осы тақырыптағы зерттеу жұмыстары.** Көпмәнді бейнелеу аппаратын машиналық аудармада пайдалану бір тілден екінші тілге аударма жасау алгоритмдерінің тиімділігін арттырады, себебі көпмәнді бейнелеулерді кестелік түрде көрсету машиналық аударма алгоритмдерінің жылдамдығын арттырып, машиналық аударманың әр кезеңіндегі көпмәнділік мәселелерін анық көрсетіп, аталған мәселені көпмәнді бейнелеулерді бірімәнді бейнелеуге түрлендіру арқылы шешеді [3].

Машиналық аудармадағы омонимия мәселесін көптеген ғалымдар шешуге тырысқан. Соның ішінде И. С. Волошина ұсынған әдіс бойынша мәтінді салаларға бөлу ұсынылады. Алдымен машиналық аударма енгізілген сөздің аудармасын «Ортақ лексика» тарауынан іздейді, бұл жағдайда аудармашыға сөздің тек жиі қолданылатын баламасын ғана алу керек болады. Егер мәтін «арнайы лексика» тарауына жатса (математика, биология, химия, физика тараулары), онда оны аударудың екі жолы бар деп қарастырады. Бірінші жолы енгізілген мәтіннің саласын бірден белгілеп, оны сол тараудан ғана іздеу, ал екінші жолы – көпмәнді сөздерді сөз тіркестері түрінде енгізу арқылы, олардың дұрыс баламасын алу.

Көпмәнді бейнелеуді формальды түрде келесідей ұсынуға болады:  $X, Y$  – дискретті кеңістік,  $P(Y)$  –  $Y$  жиынының барлық ішкі жиындарының тізбегі болсын. Сонда,  $X$ -тен  $Y$ -ке болатын  $F$  көпмәнді бейнелеуі –  $x$   $X$  әрбір нүктесіне  $x$  нүктесінің бейнесі деп аталатын бос емес  $F(x)$   $Y$  жиынын салыстыратын сәйкестік болып табылады, яғни  $F: X \rightarrow P(Y)$ . Бұл бейнелеуді **м-бейнелеу** деп атаймыз ( $X$ -тен  $Y$ -ке).

Көпмәнді бейнелеуді бірімді бейнелеуге түрлендіру үшін  $X$  жиынына  $T$  параметрлердің қосымша жиынын енгіземіз:

$$F: X \times T \rightarrow Y, \quad (1.1)$$

Сол кезде  $F$  көпмәнді бейнелеуін бірімді бейнелеулер тізбегіне түрлендіруге мүмкіндік туады:

$$\{f_i: X \rightarrow Y\}, f_i(x) \in F(x). \quad (1.2)$$

Қазақ-орыс машиналық аудармасындағы көпмәнді бейнелеу үлгісін қарастырайық. Төменде көпмәнді бейнелеу кестелерінің сегменттері көрсетілген [4].

1.1-кесте – Орыс тіліндегі зат есім қосымшаларының көпмәнді бейнелеу сегментінің кестесі [4]

Окончание	Число	Падеж	Склонение
а,я	1	1	1
ы, и	1	2	1
е	1	3	1
у, ю	1	4	1
ой, ей, ей	1	5	1
е	1	6	1
ы, и	2	1	1
-, ей, ь	2	2	1
ам, ям	2	3	1
ы, ей, и	2	4	1
	...		

Қазақ-орыс машиналық аударма келесі ретте жұмыс істейді: алдымен кіріске келіп түскен сөзге морфологиялық талдау жасалады, соның негізінде енгізілген сөзді мақсат тілге аударатын морфологиялық синтез жасалады, келесі кезектес синтаксистік талдау және соған сәйкес синтаксистік синтез жасалады.

**Қазақ-орыс машиналық аудармасында енгізілген сөздің негізі мен қосымшасын іздеу алгоритмі.** Бағдарламаның мәліметтер қорында аударылмалы сөздердің түбірлері 8 кестеде сөз таптарына қарай бөлініп орналасқан. Осыған дейін іздеу әдісі келесі түрде жұмыс істеді:

Біріншіден, сөздің алғашқы әрпін барлық кестелерден іздеп, дұрыс нұсқасы табылған жағдайда, сөздің екінші бөлігін жалғаулар кестесінен іздеп, оның атрибуттарын анықтады.

Ал табылмаған жағдайда бірінші бөлігін бір әріпке ұзартып, дәл осы операция қайтадан орындалады. Осылай қажетті нәтиже алынғанша қайталана береді.

Тексеріс жүзінде аударма жылдамдығы күрт азайып кеткені байқалды. Оның негізгі себебі әрбір сөзді 8 кестеден айналдыра іздеген өте тиімсіз. Сол себепті аудармаға қажетті мәліметтерді алудың жолын өзгерту қажет болды. Жаңа әдіс бізді күткен нәтижемізге жеткізе алды.

Жаңа іздеу алгоритмі бойынша бірінші қадамда сөздің екінші әріпінен басталып соңына дейін созылатын бөлігін жалғаулар кестесінен іздейді. Табылмаған жағдайда оң жақ бөлікті бір әріпке қысқартып отырады. Түбінде жалғау кестеден табылған кезде, сөздің бірінші бөлігі, яғни түбірі жалғаудың параметрлерінде көрсетілген сөз табының кестесінен ізделеді.

Келесі кезектегі мәселе жалғаудың қай сөз табына жалғанғандығы анықталғаннан кейін, сөздің негізін қай кестеден іздеу керектігі болып табылады. Осылайша түбірі анықталса да, анықталмаса да жалғауды қысқарта береді және мүмкін болғанша жалғаудың ең қысқа нұсқасын анықтайды.

**Қазақ-орыс машиналық аудармасының морфологиялық талдауы.** Морфологиялық талдау машиналық аударма бағдарламасының негізі болып табылады. Кез келген тілде сөздің морфологиялық құрамы болады, яғни біз сөзді аудару барысында оның құрылымын толығымен көре

аламыз. Кез келген сөздің құрылымында түбір мен қосымшасы тұрады, келесі алгоритмде қазақ-орыс машиналық аударма бағдарламасындағы сөздің морфологиялық талдауын көре аласыз:

$$A_1 = \{a_1, a_2, \dots, a_{k-1}\} \quad (1.3)$$

$A_1$  – сөйлем;  $a_1, a_2, \dots, a_{k-1}$  – сөздер.

Морфологиялық талдау барысында қазақ тілінде енгізілген сөздің негізі мен қосымшасы ажыратылады. Яғни сіз көріп тұрғандай, ең алдымен бағдарлама сөйлемді сөздерге бөледі, одан кейін толығымен сөзді алып, түбірді – part1 деп алады оны деректер қорында орналасқан 8 кестеден іздейді, ал жалғауды – part2 ретінде қарастырады, оның бастапқы мәні бос болады. (1.3) формула бойынша түбірді кестеден таппаса сөзді әр іздеу сайын бір әріпке қысқартып отырады, сәкесінше part2 бір әріпке қосып отырады.

Part1 = {"адамдар", "адамда", "адамд", "адам"} – сөз табылды → "человек"

Part2 = {"", "р", "ар", "дар"} – жалғау табылды → "и"

Нәтижесінде кіріске келген "адамдар" сөзі "люди" деп аударылады.

**Морфологиялық талдау барысында көпмәнді бейнелеу аппаратын пайдалану.** Қазақ тілінде көпмәнді қосымшалар жиі кездеседі, осы мәселені шешу үшін көпмәнді бейнелеу аппаратын пайдалану машиналық аударма жұмысының сапасын жақсартуда тиімді болып табылады. Мысалы, кіріске қазақ тіліндегі «келді» сөзі келіп түсетін болса, машина алдымен оның негізін («кел») табады және оның орыс тіліндегі баламасын алады, содан кейін қосымшасының орыс тіліндегі баламасын алу барысында табыс септігінің жалғауы «ді»-ні алу керек пе немесе өткен шақ жұрнағы «ді»-ні алу керек пе деген қиын жағдайға тап болады. Осы мәселені шешу барысында қазақ тіліндегі қосымшалар сөздікке кестелік бейнелеу түрінде енгізілді (1-сурет).

RecNo	id	zhalgau	s_e	koptik	septik	taueldik	jiktik	shak	bolymsyz
1	217	ді	2	1	0	0	3	6	<null>
2	754	ді	1	1	4	0	0	0	<null>
3	793	ді	2	2	0	0	3	6	<null>

1-сурет – Қазақ тіліндегі көпмәнді «ді» қосымшасы

**Қазақ-орыс машиналық аудармасының морфологиялық синтезін жүргізу.** Морфологиялық синтез енгізілген сөздің орыс тіліндегі аудармасы түбірінің атрибуттары мен қазақ тіліндегі қосымшалар атрибуттары негізінде жасалады.

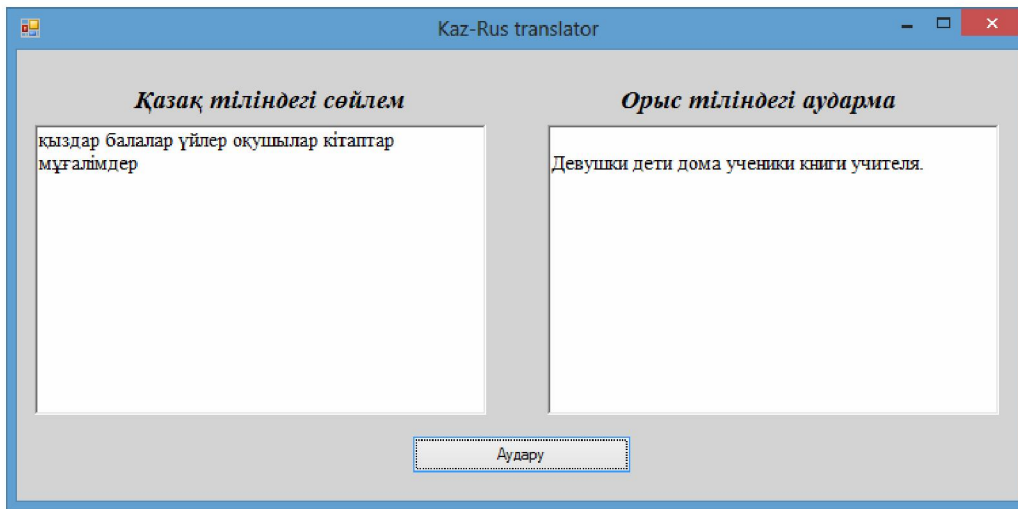
**Қазақ-орыс машиналық аудармасында көптік жалғауларын аудару алгоритмдері.** Қазақ тіліндегі көптік жалғау орыс тіліндегі «число» категориясына сәйкес келеді. Егер морфологиялық талдау барысында сөзге көптік жалғауы жалғанғаны белгілі болса, онда оның орыс тіліндегі аудармасына «множественное число» тағайындалады. Көптік формасы белгілі бір ережеге бағынбайтын, ерекше формадағы сөздер деректер базасындағы зат есім кестесіне енгізілді.

Сөзге көптік жалғауы жалғанғаны белгілі болған кезде деректер базасынан сол сөздің көптік бағаны тексеріледі, егер оның мәні 0-ге тең емес болса, онда тұрған мән алынады, мысалы, *бала – ребенок, балалар – дети*. Кері жағдайда көптік жалғауы келесі тәртіптер бойынша жалғанады:

Егер енгізілген сөздің орыс тіліндегі аудармасы «мужской родқа» тиісті болса және соңғы әрпі  $pb[] = \{ 'б', 'в', 'д', 'з', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф' \}$  әріптерінің біріне тең болса, онда сөз аудармасына «ы» жалғауы жалғанады (мысалы, *қылыштар => кинжалы*). Егер аударманың соңғы әрпі  $pb[] = \{ 'г', 'ж', 'й', 'ш', 'щ', 'ч' \}$  әріптерінің біріне тең болса, онда аударма соңына «и» қосымшасы жалғанады (мысалы, *гараждар => гаражи*) [4].

Егер сөз аудармасы «женский родқа» тиісті болса және соңғы әрпі «а», ал соңғы әріптің алдында  $pb[] = \{ 'б', 'в', 'д', 'з', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф' \}$  жиынындағы әріптерінің бірі болса, онда сөз түбірінің орыс тіліндегі аудармасына «ы» жалғауы жалғанады (мысалы, *парталар => парты*), ал егер соңғы әрпі «я» болса немесе «а» болып, соңғы әріптің алдында  $pb[] = \{ 'г', 'ж', 'й', 'х', 'ш', 'ч', 'щ', 'я' \}$  жиынындағы әріптердің бірі болса, онда сөздің орыс тіліндегі аудармасына «и» қосымшасы тағайындалады (мысалы, *жолдар => дороги*) [4].

Егер енгізілген сөздің аудармасы орыс тіліндегі «средний родқа» тиісті болса және соңғы екі әрпі «ие», «ре», «ле» болса, онда соңғы әріп «я» болып өзгереді, қалған жағдайлардың барлығында «а» қосымшасы жалғанады (мысалы, *терезелер => окна, теңіздер => моря*) [4].



2-сурет – Қазақ-орыс машиналық аудармасы бағдарламасының көпше түрдегі сөздерді аударуы

**Қазақ-орыс машиналық аудармасында септік жалғауларын аудару алгоритмдері.** Қазақ тіліндегі қосымшаның тағы бір түрі септік жалғауы болып табылады. Морфологиялық генератор жұмысын жеңілдету мақсатында орыс тіліндегі септік жалғаулары ДБ-да сәйкесінше орыс тіліндегі үш септеу бойынша кесте түрінде жинақталды. Бұл кестелерде орналасқан мәндер қазақ тіліндегі сөз атрибуттары бойынша шақырылатын болады. Ендігі кезекте жоғарыда келтірілген кестені қолдану арқылы қазақ тіліндегі септік жалғаулы сөздің орыс тіліне аударудың алгоритмдерін қарастырайық (сөздер жекеше түрде берілген):

**Ілік септігі.** 1-септеу. Ілік септігі орыс тіліндегі «родительный падежге» сәйкес келеді. Егер сөз аудармасының соңғы әрпі «а», ал соңғы әріптің алдындағы әріп  $pb[] = \{ 'б', 'в', 'д', 'з', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф', 'ц' \}$  әріптерінің біріне тең болса, сөз аудармасына «ы» қосымшасы жалғанады, мысалы *көліктің => машины*. Егер сөз аудармасының соңғы әрпі «я» болса, онда аударманың соңғы әрпі «и» болып өзгертіледі, мысалы *жердің => земли* [4].

2-септеу. Егер сөз аудармасы «мужской родқа» тиісті болса, онда «родительный падежде» сөз соңына «а» қосымшасы жалғанады, мысалы *үйдің => дома*. Егер сөз аудармасы «мужской родқа» тиісті болса және соңғы әріп «ь» болса, онда сөздің соңғы әрпі «я» әрпімен алмастырылады, мысалы *жылқының => коня*. Егер сөз аудармасы «средний род» категориясына тиісті болса, онда морфологиялық синтез жасау барысында сөздің соңғы әрпі «а» жалғауымен алмастырылады, мысалы *күннің => солнца* [4].

3-септеу. Егер сөз аудармасы «женский род» категориясына тиісті болса және оның соңғы әрпі «ь» болса, онда «родительный падежде» ол сөз аудармасының соңғы әрпі «и» жалғауымен алмастырылады, *жолдың => пути* [4].

**Барыс септігі.** Орыс тіліндегі «дательный падежге» сәйкес келеді. 1-септеу. Егер сөз аудармасы «дательный падежге» тиісті болса, сөз аудармасына «е» қосымшасы жалғанады және сөз алдына «к» предлогы жалғанады, мысалы *көлікке => к машине* [4].

2-септеу. Егер сөз аудармасы «мужской род» немесе «средний родқа» тиісті болса, онда аударманың қосымшасы «у» қосымшасымен алмастырылады және сөз алдына «к» предлогы жалғанады, мысалы *сыныпқа => к классу, күнге => солнцу*. Егер сөз аудармасы «мужской родқа» тиісті болса және оның соңғы әрпі «ь» болса, онда сөзге «и» қосымшасы жалғанады [4].

3-септеу. Егер сөз аудармасы «женский род» категориясына тиісті болса және оның соңғы әрпі «ь» болса, онда қосымша «и» жалғауымен алмастырылады және сөз алдына «к» предлогы жалғанады, мысалы *жолға => к пути* [4].

**Табыс септігі.** Орыс тіліндегі «винительный падежге» сәйкес келеді. 1-септеу. Егер сөз аудармасы «винительный падежде» тұрса және қосымша «а», ал соңғы әріптің алдындағы әріп  $pb[] = \{ 'б', 'в', 'г', 'д', 'ж', 'з', 'к', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф', 'ч', 'ш', 'щ', 'х', 'ц' \}$  әріптерінің біріне тең болса, сөз аудармасына «у» жалғауы жалғанады, мысалы *көлікті =>машину*. Егер сөздің соңғы әрпі «я» болса, онда аудару барысында сөздің қосымшасы «ю» болып өзгереді, мысалы *жерді =>землю* [4].

2-септеу. Егер сөз аудармасы мужской родқа тиісті болып, оның соңғы әрпі «ь» болса және сөз аудармасы средний родқа тиісті болып, оның соңғы екі әрпі «ие», «ре», «ле» болса, сөз қосымшасы «я»-ға ауысады, қалған жағдайларда «а» жалғауы жалғанады, мысалы *терезені =>окна, теңізді =>моря* [4].

3-септеу. Егер сөз аудармасы орыс тіліндегі үшінші септеуге сәйкес келсе, онда «винительный падежде» сөз түбірінің өзі шығарылады, мысалы *жолды =>путь* [4].

**Жатыс септігі.** Орыс тіліндегі «предложный падежге» сәйкес келеді және аударманың алдына «на» немесе «в» предлогтарының бірі қойылуы тиіс. Егер сөз «мужской род», «женский род» немесе «средний родқа» тиісті болса, онда сөз түбірінің аудармасына «е» жалғауы жалғанады да, сөз алдына «в» немесе «на» предлогы қойылады. Мысалы *көлікте =>в машине, үйде=>в доме* [4].

**Шығыс септігі.** Орыс тіліне аударылу барысында «родительный падежге» сәйкес аударылады, тек қана аударманың алдына «из» немесе «от» предлогы жалғанады.

1-септеу. Егер сөз аудармасының соңғы әрпі «а», ал соңғы әріптің алдындағы әріп  $pb[] = \{ 'б', 'в', 'д', 'з', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф', 'ц' \}$  әріптерінің біріне тең болса, сөз аудармасына «ы» қосымшасы жалғанады және «из/от» предлогы қойылады, мысалы *көліктен => из машины*. Егер сөз аудармасының соңғы әрпі «я» болса, онда аударманың соңғы әрпі «ю» болып өзгертіледі және «из/от» предлогы қойылады, мысалы *автоматтандырудан=> из автоматизации* [4].

2-септеу. Егер сөз аудармасы «мужской родқа» тиісті болса, онда сөз соңына «а» қосымшасы жалғанады және сөз алдына «из/от» предлогы қойылады, мысалы *үйден=>из дома*. Егер сөз аудармасы «мужской родқа» тиісті болса және соңғы әріп «ь» болса, онда сөздің соңғы әрпі «я» әрпімен алмастырылады және «из/от» предлогы қойылады, мысалы *ажыратқыштан => из выключателя*. Егер сөз аудармасы «средний род» категориясына тиісті болса, онда морфологиялық синтез жасау барысында сөздің соңғы әрпі «а» жалғауымен алмастырылады және «из/от» предлогы қойылады, мысалы *терезеден => из окна* [4].

3-септеу. Егер сөз аудармасы «женский род» категориясына тиісті болса және оның соңғы әрпі «ь» болса, онда сөз аудармасының соңғы әрпі «и» жалғауымен алмастырылады және «из/от» предлогы қойылады, *жолдан =>из пути* [4].

**Көмектес септік.** Орыс тіліндегі «творительный падежге» сәйкес келеді.

1-септеу. Егер орыс тіліндегі сөз «творительный падеж» формасында тұрса және оның соңғы әрпі «а», ал соңғы әріптің алдындағы әріп  $pb[] = \{ 'б', 'в', 'г', 'д', 'ж', 'з', 'к', 'л', 'м', 'н', 'п', 'р', 'с', 'т', 'ф', 'ч', 'ш', 'щ', 'х' \}$  әріптерінің біріне тең болса, онда сөзге «ой» жалғауы жалғанады, мысалы *көлікпен =>машинной*. Егер сөздің соңғы әрпі «ь» болса, онда морфологиялық синтез барысында сөзге «ей» жалғауы жалғанады, мысалы *жермен =>землей* [4].

2-септеу. Егер сөз аудармасы «мужской родқа» тиісті болса, синтез жүргізу барысында сөз түбіріне «ом» жалғауы жалғанады, мысалы *үймен =>домом*. Егер сөз аудармасы «мужской родқа» тиісті болса және оның соңғы әрпі «ь» болса немесе сөз аудармасы «средний родқа» тиісті болса, онда сөз түбіріне «ем» жалғауы жалғанады, мысалы *күнмен =>солнцем* [4].

3-септеу. Егер сөз аудармасы орыс тіліндегі үшінші склонениеге сәйкес келсе, онда бұл септікте сөз соңына «ю» жалғауы жалғанады, мысалы *жолмен =>путью* [4].

Септік атрибуттары жоғарыда келтірілген кестеге сәйкес орыс тіліне аударылады. Жоғарыда көрсетілген алгоритмдер жекеше түрдегі зат есімдерді септеуге арналған. Егер морфологиялық синтез кірісіне көпше түрдегі зат есім кіретін болса, онда алдымен сөз аудармасы көпше түрге ауыстырылады және тек содан кейін септеледі. Көпше түрдегі зат есімдердің септік жалғаулары да деректер базасына енгізілген [4].

**Қазақ-орыс машиналық аудармасының синтаксистік талдауы.** Машиналық аударманың сапалы болуының тағы бір шарты – сөйлемге дұрыс жасалған синтаксистік талдау мен сол талдауға байланысты құрастырылған синтаксистік синтездің болуы. Синтаксистік талдау барысында

қазақ тілінде енгізілген сөздер сөйлем мүшелеріне жіктеледі. Тілдегі сөйлемді құрауыштар арқылы формальды грамматика түрінде көрсететін болсақ:

- сөйлем (sentence-S)
- бастауыш тобы (noun phrase-NP)
- баяндауыш тобы (verb phrase-VP)
- пысықтауыш тобы (adverbial phrase-Adv.P)
- анықтауыш тобы (adjectival phrase-Adj.P)
- толықтауыш тобы (addition phrase-AdP)

Қазақ тіліндегі сөйлем құрылымын келесі формальды грамматика түрінде көрсетуге болады:

<сөйлем> ::= <бастауыш> <баяндауыш> <толықтауыш> <анықтауыш>

<пысықтауыш>

<бастауыш> ::= <атау септігіндегі зат есім> | <есімдік>

<баяндауыш> ::= <етістік> | <құрделі етістік>

<толықтауыш> ::= <табыс септігіндегі зат есім> | <табыс септігіндегі есімдік> | <көмектес септігіндегі зат есім>

<анықтауыш> ::= <сын есім> | <ілік септігіндегі зат есім>

<пысықтауыш> ::= <үстеу> | <барыс септігіндегі зат есім> | <жатыс септігіндегі зат есім> | <шығыс септігіндегі зат есім>

Қазақ-орыс машиналық аудармасының жұмысын жеңілдету мақсатында қазақ және орыс тіліндегі сөйлем мүшелерінің сәйкестіктері орнатылып, оларға арнайы нөмірлер тағайындалды. Осы нөмірлер арқылы қазақ және орыс тілдеріндегі сөйлем құрылымын тұрғызуға болады.

2-кесте – Қазақ және орыс тілдеріндегі сөйлем мүшелерінің сәйкестігі

Атрибут	Қазақ тіліндегі сөйлем мүшесі	Орыс тіліндегі сөйлем мүшесі
1	Бастауыш	Подлежащее
2	Баяндауыш	Сказуемое
3	Толықтауыш	Дополнение
4	Анықтауыш	Определение
5	Пысықтауыш	Обстоятельство

**Қазақ-орыс машиналық аудармасының синтаксистік синтезі.** Синтаксистік синтез – синтаксистік талдау жасалған сөйлемдегі сөйлем мүшелерін дұрыс ретпен орналастыру. Синтаксистік синтез барысында белгілі бір құрылымдағы қазақ тілінде енгізілген сөйлемнің орыс тіліндегі дұрыс құрылымы тағайындалады. Сөйлем мүшелерінің сөйлемдегі орындарының реті келесі сәйкестікте болады, теңдіктің сол жағында қазақ тіліндегі сөйлем құрылымы, ал теңдіктің оң жағында қазақ тіліндегі сөйлем құрылымына сәйкес келетін орыс тіліндегі сөйлем құрылымы орналасқан:

[1][2]=[1][2]

[1][4][4][3][2]=[1][4][2][4][3]

[3][2]=[3][2]

[1][4][3][4][2]=[1][4][3][4][2]

[1][3][2]=[1][2][3]

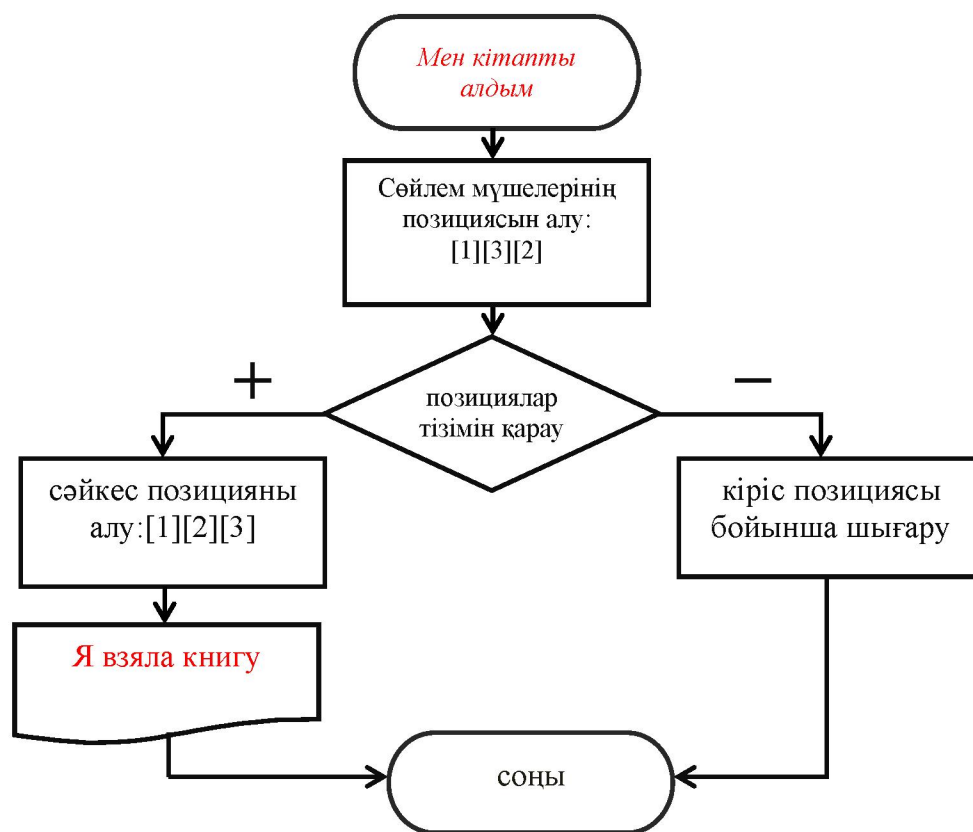
[1][3][4][4][2]=[1][2][3][4][4]

[1][4][2]=[1][2][4]

[1][4][4][3][2]=[1][2][4][4][3]

Бұл жерде қазақ тіліндегі алты сөзге дейінгі жиі кездесетін сөйлем құрылымдары жинақталған, егер машина енгізілген құрылымды бұл тізімнен таппаса, онда шығысындағы сөйлем кіріске келіп түскен позиция бойынша шығарылады (орыс тілінде сөйлем мүшелерінің реті орнықты болуы шарт емес). Қазақ тіліндегі сөйлемге синтаксистік талдау жасалып, олардың сөйлемдегі орналасу реттері анықталғаннан кейін жоғарыда келтірілген сәйкестік бойынша орыс тіліндегі сөйлем құрастырылады. Мысалы, қазақ тіліндегі [1][3][2] құрылымына сәйкес келетін «Мен кітапты алдым» сөйлемі енгізілетін болса, ондағы сөздердің қай сөз таптарына жататындықтары анықталады. Ендігі кезекте ол сөз таптарының қай сөйлем мүшесіне сәйкес келетіндіктері жоғарыда келтірілген формальды грамматика арқылы анықталады (3-сурет).

Кіріс: «Мен кітапты алдым», құрылымы [1][3][2]



3-сурет – Қазақ-орыс машиналық аудармасының синтаксистік синтез алгоритмі

**Қазақ-орыс машиналық аудармасының жұмыс істеу нәтижелері.** Қазіргі таңдағы қазақша-орысша машиналық аудармасының жұмыс нәтижесін келесі аударылған сөйлемдерден көруге болады:

<i>Қазақ тіліндегі сөйлем</i>	<i>Орыс тіліндегі аударма</i>
<p>мен ертең ойнамақпын. сен ертең ойнамақсың. ол ертең ойнамақ. біз ертең ойнамақпыз. сендер ертең ойнамақсыңдар. мен кеше келдім. мен кеше ойнағанмын. біз кеше келдік. біз кеше ойнадық. сен кеше ойнадың. сендер кеше ойнадыңдар. ол кеше келді. ол кеше ойнады.</p>	<p>Я завтра собираюсь играть. Ты завтра собираешься играть. Он завтра собирается играть. Мы завтра собираемся играть. Вы завтра собираетесь играть. Я вчера приходил. Я вчера играл. Мы вчера приходили. Мы вчера играли. Ты вчера играл. Вы вчера играли. Он вчера приходил. Он вчера играл.</p>
<p>Аудару</p>	

4-сурет – Қазақ-орыс машиналық аудармасының өткен шақ пен мақсатты келер шақты аудару мысалы

<b>Қазақ тіліндегі сөйлем</b>	<b>Орыс тіліндегі аударма</b>
<p>мен ертең ойнаймын. сен ертең ойнайсың. ол ертең ойнайды. біз ертең ойнаймыз. сендер ертең ойнайсындар.</p> <p>мен ертең ойнармын. сен ертең ойнарсың. ол ертең ойнар. біз ертең ойнармыз. сендер ертең ойнарсындар.</p>	<p>Я завтра буду играть. Ты завтра будешь играть. Он завтра будет играть. Мы завтра будем играть. Вы завтра будете играть.</p> <p>Я завтра буду играть, наверно. Ты завтра будешь играть, наверно. Он завтра будет играть, наверно. Мы завтра будем играть, наверно. Вы завтра будете играть, наверно.</p>
<div style="border: 1px dashed black; padding: 2px; display: inline-block;">Аудару</div>	

5-сурет – Қазақ-орыс машиналық аудармасының ауыспалы және болжалды келер шақты аудару мысалы

**Қазақ-орыс машиналық аудармасының келешектегі жұмысы.** Қорытындылай келе айтарымыз – қазақ-орыс машиналық аудармасы қазіргі таңда қарапайым сөйлемдерге сапалы аударма жасай алады, соның ішінде зат есімге септік, көптік, тәуелдік жалғауларын жалғау, сын есім мен зат есім арасындағы тәуелділікті орнату, осы шақ сөйлемдерін аудару. Соңғы атқарылған жұмыстар нәтижесінде қазақ-орыс машиналық аударма бағдарламасына өткен шақ және келер шақ алгоритмдері қосылып, аталған шақтағы сөйлемдер де сапалы түрде аударылуда. Келесі кезекте атқарылуы тиіс тапсырмалар – зат есім фразалары мен етістік фразаларын аудару алгоритмдерін қосу және бағдарламаның жұмыс тиімділігін арттыру.

#### ӘДЕБИЕТ

- 1 Введение в теорию многозначных отображений: Составитель Б. Д. Гельман. – Воронеж, 2003.
- 2 Tomasz Kaczynski, Multivalued Maps As a Tool in Modeling and Rigorous Numerics. Departement de mathematiques. – Universite de Sherbrooke, 2008.
- 3 Тукеев У.А., Жуманов Ж.М., Сапакова С.З. Метод многозначных отображений для машинного перевода с одного естественного языка на другой. – 2013.
- 4 Баскаков Н.А., Хасенова А.К., Исенгалиева В.А., Қордабаев Т.Р. Қазақ және орыс тілдерінің салыстырмалы грамматикасы. – Алматы: Наука, 1966.

#### REFERENCES

- 1 Vvedenie v teoriju mnogoznachnyh otobrazhenij: Sostavitel' B. D. Gel'man. Voronezh, 2003.
- 2 Tomasz Kaczynski, Multivalued Maps As a Tool in Modeling and Rigorous Numerics. Departement de mathematiques. Universite de Sherbrooke, 2008.
- 3 Tukeev U.A., Zhumanov Zh.M., Sapakova S.Z. Metod mnogoznachnyh otobrazhenij dlja mashinogo perevoda s odnogo estestvennogo jazyka na drugoj. 2013.
- 4 Baskakov N.A., Hasenova A.K., Isengalieva V.A., Kordabaev T.R. Қазақ және орыс тилдеринин salystymnaly grammatikasy. Almaty: Nauka, 1966.

#### Резюме

У. А. Тукеев, М. А. Болатбек, М. А. Ильжанов, А. Х. Разахов

(Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан)

#### КАЗАХСКО-РУССКИЙ МАШИННЫЙ ПЕРЕВОД НА ОСНОВЕ АППАРАТА МНОГОЗНАЧНЫХ ОТОБРАЖЕНИЙ

*Цель работы:* разработка моделей, алгоритмов и программы казахско-русского машинного перевода на основе аппарата многозначный отображений, обеспечить отечественных пользователей качественной программой машинного перевода.

*Результаты:* была разработана программа, выполняющая перевод слов, словосочетаний и простых предложений с казахского на русский язык.



*Область применения:* предложенной программой может пользоваться любой человек, нуждающийся в помощи переводчика.

*Научная новизна:* поиск корня и окончаний введенного слова выполняется с двух сторон, соответствия окончаний казахского и русского языков выполняется с помощью запрашивания соответствующих суффиксов из базы данных, в базе данных имеется более 12 000 слов, разделенных по частям речи.

**Ключевые слова:** машинный перевод, анализатор машинного перевода, генератор машинного перевода, морфологический анализ, морфологический генератор, алгоритмы машинного перевода, синтаксический анализ, синтаксический генератор, аппарат многозначный отображений.

### Summary

*U. A. Tukeev, M. A. Bolatbek, M. A. Il'zhanov, A. Kh. Razakhov*

(Al-Farabi Kazakh national university, Almaty, Kazakhstan)

#### KAZAKH-RUSSIAN MACHINE TRANSLATION BASED ON THE UNIT-VALUED MAPPINGS

*Purpose of work:* the system analysis of existing domestic machine translation systems, development of models, algorithms and the program of the analyser and generator of the Kazakh-Russian machine translation based on machine set-valued mapping.

*Results:* the program which is carrying out the translation of words, phrases and simple sentences from Kazakh into Russian was developed.

*Scope:* any person needing the help of the translator can use the offered program.

*Scientific novelty:* search of a root and the terminations of the entered word is carried out from two parties, compliances of the terminations of the Kazakh and Russian languages is carried out by means of requesting of the corresponding suffixes from a database, in a database is available more than 12 000 words divided in parts to speech.

**Keywords:** machine translation, analyser of machine translation, generator of machine translation, morphologic parse, morphologic generator, algorithms of machine translation, syntactic analyser, syntactic generator, set-valued mapping.