

*М. А. БОЛАТБЕК<sup>1</sup>, К. Д. БАЙСЫЛБАЕВА<sup>2</sup>*

<sup>1</sup>әл-Фараби атындағы Қазақ ұлттық университеті, Алматы, Қазақстан,

<sup>2</sup>Алматы Экономика және статистика академиясы, Алматы, Қазақстан)

## **ҚАЗАҚ-ОРЫС ЖӘНЕ КЕРІ БАҒЫТТАҒЫ МАШИНАЛЫҚ АУДАРМАНЫҢ ГРАММАТИКАЛЫҚ СӘЙКЕСТІКТЕРІ**

**Аннотация.** Зерттеу жұмысының мақсаты қазақша-орысша және орысша-қазақша машиналық аударма анализаторы мен генераторын құруға қажетті алгоритмдер мен үлгілерді қалыптастыру, машиналық аударма бағдарламаларын құру болып табылады. Жұмыс нәтижесі ретінде қазақ тіліндегі сөздерді, сөз тіркестері мен қарапайым сөйлемді орыс тіліне және кері бағытта сапалы түрде аударатын машиналық аударма бағдарламасы құрастырылды. Жұмыс нәтижесінің қолданылу аймағы – кез келген саладағы мәтіндерді бастапқы тілден (қазақ, орыс) мақсат тіліне (орыс, қазақ) аударғысы келетін кез келген қолданушы. Зерттеудің ғылыми жаңалықтары қазақ және орыс тілдерінде грамматикалық ерекшеліктер мен ұқсастықтар көрсеткіштік кесте түрінде жинақталғандығы болып табылады, оларды бағдарламада қолдану барысында тек сұраныс жасау жеткілікті.

**Тірек сөздер:** машиналық аударма, машиналық аударма анализаторы, машиналық аударма генераторы.

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

**Keywords:** machine translation, analyser of machine translation, generator of machine translation.

Қазақ тілі түркі тілдерінің қатарына жататын және оның осы топтағы тілдерге қатысты өзіндік ерекшеліктерін сақтаған тіл болып табылады. Қазақ тілі типі мен морфологиялық құрылымы жағынан агглюнативті тілдер қатарына жатады, сондықтан қазақ тіліндегі жаңа сөздер оның түбіріне немесе негізіне аффикстердің, грамматикалық суффикстердің (жұрнақтар) және қосымшалардың (жалғаулар) жалғануы арқылы құралады.

Орыс тілі флективті тілдер қатарына жатады. Орыс тіліндегі сөздер түбірге род, түр, септік сияқты үш грамматикалық категория сипаттарын флексия түрінде байланыстыратын сөз түрлендіруші аффикстерді жалғау арқылы құралады.

**Қазақ тілі грамматикасының мәтінді қазақ тілінен орыс тіліне аудару барысында жеңілдік тудыратын тұстары:**

– Көптеген түркі тілдеріндегі сияқты қазақ тілінде де сөз тіркестері мен сөйлемдегі сөздердің орны, сөздер арасындағы байланыс қатаң түрде анықталады және тұрақты болып келеді [1]:

а. баяндауыш әрдайым сөйлемнің соңында орналасады: *Мен университетте оқимын (Я учусь в университете);*

б. бастауыш әрдайым сөйлемнің бас жағында орналасады: *Сен Университетте оқисың (Ты учишься в университете);*

в. анықтауыш анықталуы тиіс сөздің алдында орналасады: *Мен Ұлттық университетте оқимын (Я учусь в Национальном университете);*

– Әртүрлі категориядағы етістікті тудыратын етістіктің негізі болып II жақтағы бұйрық рай формасындағы етістік табылады: *ойна /играй/ -ойна-й-ды /играет/; ойна-ды /играл/; ойна-мақ, /будет играть/;*

– Орыс тіліндегі «предлог» және «приставка» орнына қазақ тілінде аффикстер мен көмекші сөздер қолданылады: *налево – солға қарай; приехал – келді; до завтра – ертеңге дейін* [2].

Жоғарыда айтып кеткендей, қазақ тіліндегі сөз құрамының тұрақты болуы, яғни алдымен сөз түбірі, одан кейін қосымшаның орналасуы сөзге морфологиялық талдау мен синтез жасау барысында жеңілдік туындатады.

**Қазақ тіліндегі мәтінді орыс тіліне аудару барысында қиындық тудыратын тұстар:**

– Қазақ тіліндегі сөзден кейін орналасатын шылаулар орыс тілінің «предлогтарына» сәйкес келеді. Алайда қазақ тілінде әр септіктің тұрақты шылаулары болатын болса, орыс тілінде бір «предлог» бірнеше септікке сәйкес келеді, шылауы бар мәтінді орыс тіліне аудару барысында қажетті «предлогтың» қайсысын таңдау мәселесі туындайды, мысалы шығыс септігіндегі «үйімнен» сөзі енгізілетін болса, оның аудармасында орыс тіліндегі «родительный падеждің» **от, до, из, у, с** «предлогтарының» қайсысын таңдауды машина дәл таба бермейді, яғни *от дома, из дома, с дома, у дома, до дома* нұсқаларының ішінен бірі таңдалуы тиіс.

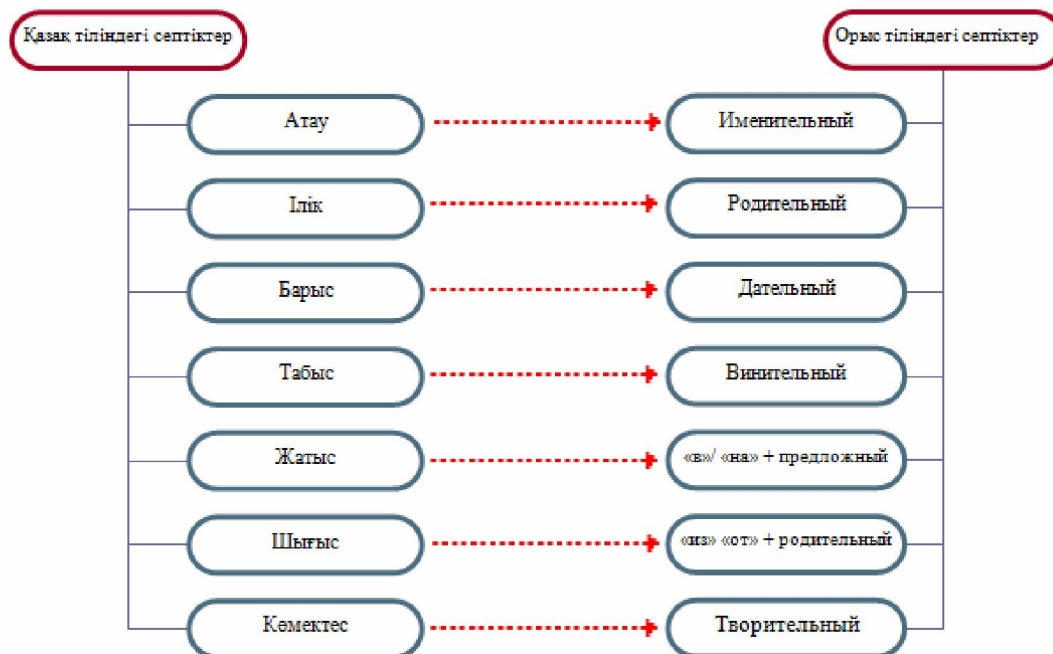
– Қазақ тілінде тәуелділікті білдіретін есімдік жасырын орналаса алады, яғни «үйімнің есігі» тіркесін көргенде біз ол үйдің бірінші жаққа тиісті екендігін түсінеміз, енді сол тіркесті орыс тіліне аударатын болсақ, тәуелділікті білдіретін есімдікті міндетті түрде қолдануымыз керек болады, яғни «*дверь моего дома*», есімдікті көрсетпей жай ғана «*дверь дома*» дейтін болсақ, мәтіннің мағынасы өзгеріп кетуі мүмкін. Тағы бір айта кететін жайт – орыс тілінде тәуелдік формасында тұрған сөзді септеу барысында оның алдында тұрған есімдік те міндетті түрде септеледі (мысалы, *менің үйімнің=> моего дома, менің үйіме => моему дому*).

– Қазақ тіліндегі етістіктің жіктік жалғаулары әр жақ үшін тұрақты болып келетін болса, яғни бірінші жақ үшін **-мын, -мін, -бын, -бін, -пын, -пін** болатын болса, орыс тіліндегі әр етістіктің өзіндік жіктелу формасы бар және көп жағдайда олар ешқандай заңдылықтарға бағынбастан өзгере береді, мысалы қазақ тіліндегі «*мен ұшамын*», «*сен ұшасың*», «*ол ұшады*» тіркестерінде етістіктің «**ұш**» түбірі сақталып, тек жіктік жалғаулары жалғанып отырса, орыс тілінде «*я лечу*», «*ты летишь*», «*он летит*» болып бастапқы сөз түбірінің «**ле**» бөлігі ғана сақталады. Орыс тілінде етістіктердің жіктелуінде ережеге бағынбайтын етістіктер көптеп кездеседі. Бұл мәселе етістіктерді аудару барысында қиындықтар тудырады.

– Қазақ тіліндегі омоним сөздердің көптеп кездесуі аудару барысында машинаға қиындық тудырады, яғни машина сөздің қай нұсқасының аудармасын алу керектігін білмей, семантика тұрғысынан сәйкес келмейтін нұсқаны алуы мүмкін.

Машиналық аударма кірісіне келіп түскен сөзге алдымен морфологиялық талдау жасалады. Морфологиялық талдау барысында сөздің негізі мен қосымшасы ажыратылады. Сөз негізі мен

қосымшасы және олардың сәйкес атрибуттары екі тілдегі деректер қорында орналасады. Қазақ тіліндегі септеулік шылаулар септелген сөздерден кейін орналасады, сол себепті қазақша-орысша машиналық аудармада кіріс сөзінен кейін тұрған септеулік шылауға қарағанда оның алдындағы сөзге назар көбірек аударылады. Соған орай қазақша-орысша машиналық аудармада септіктердің келесі сәйкестігін пайдаланған жөн.



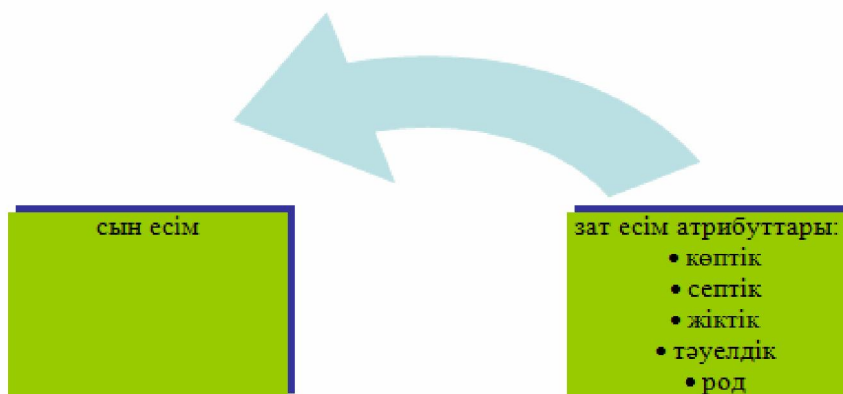
1-сурет – Қазақша-орысша машиналық аудармасындағы септіктер сәйкестігі

Орыс тілінде морфологиялық омонимдер өте жиі кездеседі. Мысалы, «*дома*» сөзіндегі «а» қосымшасын қарастырсақ, машинаның оны аудару барысында көптік жалғауының (*дома – үйлер*), жатыс септігінің жалғауының (*дома – үйде*) әлде шығыс септігінің жалғауының (*из дома – үйден*) қайсысын таңдау керек деген мәселе туындайды. Аталған мәселені шешу барысында орысша-қазақша машиналық аудармасы үшін септіктердің келесі сәйкестіктері орнатылды (1-кесте).

1-кесте – Екі тіл септіктерінің сәйкестігі

Орыс тіліндегі септік түрлері	Қазақ тіліндегі септік түрлері
Именительный падеж	Атау септік
Родительный падеж	Ілік септік
Дательный падеж	Барыс септік
Винительный падеж	Табыс септік
Творительный падеж	Көмектес септік
Родительный падеж + из	Шығыс септік
Родительный падеж + от	Шығыс септік
Предложный падеж + на	Жатыс септік
Предложный падеж + в	Жатыс септік

Қазақ тілінде зат есім мен сын есім арасында қандай да бір грамматикалық тәуелділік болмайды, яғни септік, көптік жалғаулары тек қана анықталып тұрған зат есімге жалғанады. Ал орыс тілінде зат есім мен сын есімнің арасында грамматикалық тәуелділіктің болуы міндетті, мысалы егер зат есім септелетін болса, онда оның алдында тұрған сын есімге де сәйкесінше септік жалғаулары жалғануы тиіс [2]. Қазақ тіліндегі сын есім орыс тіліне өзі анықтап тұрған зат есімнің орыс тіліндегі «родына», көптік формасына және септігіне байланысты аударылады. Қазақша-орысша машиналық аударма генераторы мәтін ішінде сын есім кездескен жағдайда оған өзі анықтап тұрған (сын есімнен кейін орналасқан) зат есімнің атрибуттарын тағайындайды, яғни оның көптік, септік жалғаулары мен «роды» өзінен кейін орналасқан зат есімнің атрибуттарына тең болады (2-сурет):



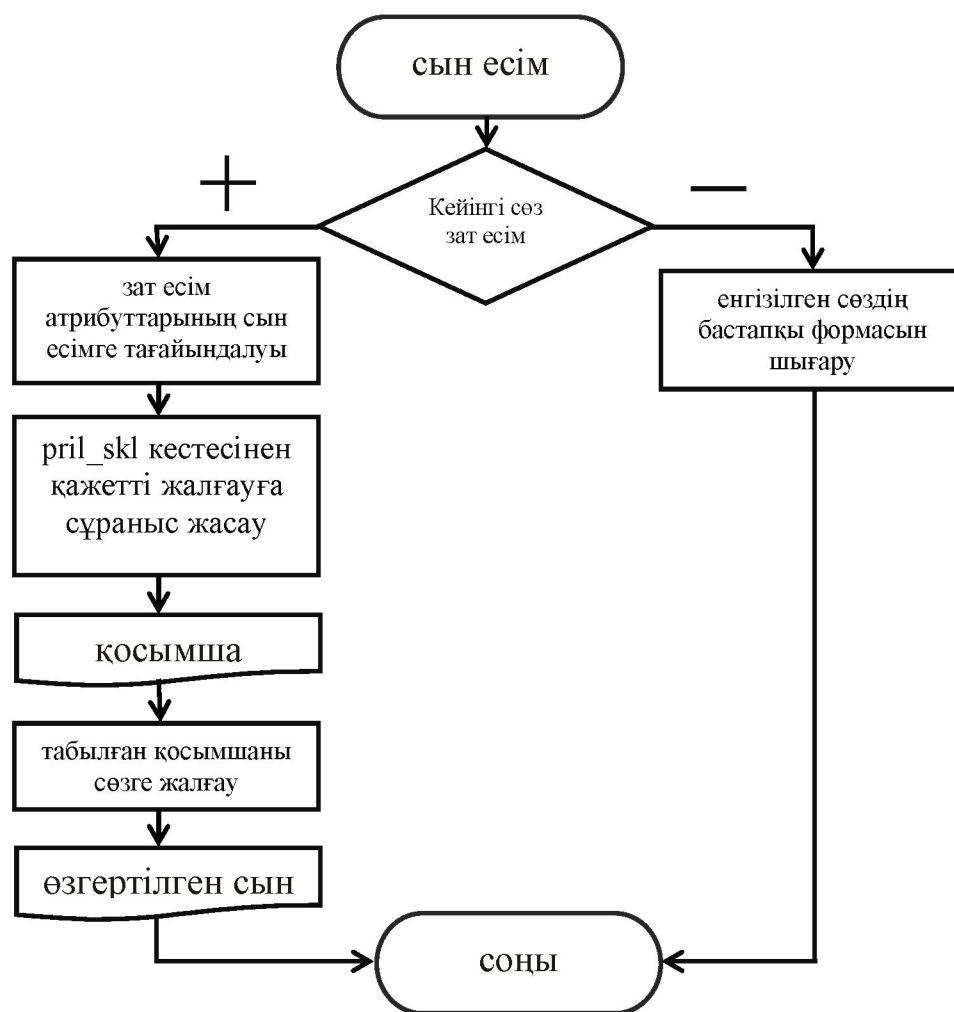
2-сурет – Кейінгі тұрған зат есім атрибутының сын есімге тағайындалуы

Орыс тіліндегі сын есімдердің де тұрақты септік жалғаулары болады. Олар бағдарламада қолданылу үшін деректер базасына жеке кесте ретінде толтырылды (3-сурет):

RecNo	padezh	rod	chislo	okonch1	okonch2
1	2	1	1	ого	его
2	3	1	1	ому	ему
3	1	1	1	ый	ий
4	4	1	1	ого	его
5	5	1	1	ым	им
6	6	1	1	ом	ем
7	0	2	1	ая	яя
8	2	2	1	ой	ей
9	3	2	1	ой	ей
10	4	2	1	ую	юю
11	5	2	1	ой	ей

3-сурет – Деректер базасындағы сын есімнің септік жалғаулары

Сын есімді аудару үшін кейінгі тұрған зат есімнің атрибуттары (род, көптік, септік) сын есімге берілгеннен кейін жоғарыдағы кестеден аталған атрибуттарға ие жалғауға сұраныс жасалады. Қажетті жалғау табылғаннан кейін олар сын есімге жалғанады, осы алгоритмді келесі түрде көрсетуге болады (4-сурет):



4-сурет – Қазақ-орыс машиналық аударма генераторының сын есімді аудару алгоритмі

Деректер базасындағы *okonch1* және *okonch2* жалғауларын таңдау сын есімнің соңғы әріптеріне байланысты жүргізіледі. Мысалы, мәтін ішінде кездесетін «қызыл кітаптың» тіркесін аудару қажет болса, алдымен сын есімнен кейін тұрған «кітап» сөзінің атрибуттары сын есімге тағайындалады. Кіріс мәліметтері:

<кітаптың>::=<кітап><тың>

<кітап>::=<rod=2,skl=1>

<тың>::=<padezh=2, chislo=1>

Сын есімге тағайындалатын атрибуттар:

<қызыл>::=<rod=2, padezh=2, chislo=1 >

Деректер базасындағы *pril\_skl* кестесінен *rod=2*, *padezh=2*, *chislo=1* болатын жалғауға сұраныс жасалады және табылған қосымша сын есімге жалғанады.

Ал орыс тілінен қазақ тіліне аударма жасау барысында келтірілген мәселе үлкен қиындық туғызбайды, себебі – қазақ тілінде «род» ұғымы болмайды, мысалы *синяя книга* – **көк** кітап, *синее небо* – **көк** аспан және *синий стол* – **көк** үстел.

Морфологиялық талдау мен генератор жұмыстары аяқталғаннан кейін синтаксистік талдау мен генератор жұмыстары басталады. Қазақ тілінде сөйлем мүшелерінің орындары көбінесе тұрақты болып келеді, сол себепті орысша-қазақша машиналық аудармада қазақ тіліндегі сөйлем мүшелерінің ретін анықтап, оларды өз орындарына орналастыру қажеттілігі туындайды. Бұл мәселені шешу үшін екі тілдегі сөйлем мүшелерінің сәйкестіктері орнатылды:

Подлежащее – Бастауыш {1}

Сказуемое – Баяндауыш {2}

Дополнение – Толықтауыш {7}  
 Определение – Анықтауыш {8}  
 Обстоятельство – Пысықтауыш {9}

2-кесте – Орыс тілі мен қазақ тілінің сөйлем мүшелерінің сәйкестігі

Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я пойду	Мен барамын
{1 2}	{1 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Читать книгу	Кітап оқу
{2 7}	{7 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я читаю книгу	Мен кітап оқып отырмын
{1 2 7}	{1 7 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Красивая девушка танцует	Шырайлы қыз билеп жүр
{4 1 2}	{4 1 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я спешу домой	Мен үйге асығып барамын
{1 2 5}	{1 5 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я пишу письмо подруге	Мен құрбыма хат жазып отырмын
{1 2 7 5}	{1 5 7 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я на больших концертах танцую	Мен үлкен концерттерде билеймін
{1 4 7 2}	{1 4 7 2}
Орыс тіліндегі сөйлем және реті	Қазақ тіліндегі сөйлем және реті
Я читаю интересную книгу в библиотеке	Мен кітапханада қызықты кітап оқып отырмын
{1 2 4 7 5}	{1 5 4 7 2}

Қазіргі таңда жоғарыда келтірілген алгоритмдер бойынша жұмыс істейтін қазақша-орысша және орысша-қазақша машиналық аударма бағдарламалары бар. Олардың көмегімен қарапайым сөздерді, сөз тіркестерін және жай сөйлемдерді аударуға болады. Болашақта екі тілдің ерекшеліктері мен ұқсастықтарын ескере отырып, күрделі сөйлемдерді, етістік фразаларын және т.б. лексикалық бірліктерді аудару алгоритмдерін машиналық аударма бағдарламасына қосу жоспарлануда.

#### ӘДЕБИЕТ

- 1 Баскаков Н.А., Хасенова А.К., Исенғалиева В.А., Қордабаев Т.Р. Қазақ және орыс тілдерінің салыстырмалы грамматикасы. – Алматы: Наука, 1966.  
 2 <http://www.testent.ru/index/0-199>

#### REFERENCES

- 1 Baskakov N.A., Hasenova A.K., Isengalieva V.A., Kordabaev T.R. Қазақ және орыс тілдерінің салыстырмалы грамматикасы. – Алматы: Наука, 1966.  
 2 <http://www.testent.ru/index/0-199>



## Резюме

*М. А. Болатбек<sup>1</sup>, К. Д. Байсылбаева<sup>2</sup>*

(<sup>1</sup>Казахский национальный университет им. аль-Фараби, Алматы, Казахстан,  
<sup>2</sup>Алматинская академия Экономики и статистики, Алматы, Казахстан)

### ГРАММАТИЧЕСКИЕ СООТВЕТСТВИЯ МАШИННОГО ПЕРЕВОДА С КАЗАХСКОГО НА РУССКИЙ ЯЗЫК И ОБРАТНО

*Цель работы:* формировать алгоритмы и модели для разработки анализатора и генератора казахско-русского и русско-казахского машинного перевода, разработка программы машинного перевода.

*Результат:* разработана программа перевода слов, словосочетаний и простых предложений с казахского на русский язык и обратно.

*Область применения:* все пользователи, нуждающиеся в переводе текста из исходного языка (казахский, русский) на целевой язык (русский, казахский)

*Новизна работы:* грамматические особенности и сходства двух языков представлены в виде индексных таблиц, при использовании их в программе достаточно сделать запрос.

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

## Summary

*M. A. Bolatbek<sup>1</sup>, K. D. Baisylbaeva<sup>2</sup>*

(<sup>1</sup>Al-Farabi Kazakh national university, Almaty, Kazakhstan,  
<sup>2</sup>Алматы Экономика және статистика академиясы, Almaty, Kazakhstan)

### GRAMMATICAL MATCHING MACHINE TRANSLATION FROM KAZAKH INTO RUSSIAN AND BACK

*Work purpose:* to form algorithms and models for development of the analyzer and the generator of the Kazakh-Russian and Russian-Kazakh machine translation, development of the program of machine translation.

*Result:* the translation program of words, phrases and simple sentences with Kazakh into Russian and back is developed.

*Scope:* all users needing the translation of the text from source language (Kazakh, Russian) into target language (Russian, Kazakh)

*Novelty of work:* grammatical features and similarities of two languages are presented in the form of index tables, when using them in the program it is enough to make request.

**Keywords:** machine translation, analyser of machine translation, generator of machine translation.

*Поступила 05.05.2014 г.*