Proceedings of the International Multiconference on
Computer Science and Information Technology, pp. 363 – 368

ISBN 978-83-60810-14-9
ISSN 1896-7094

Set of Active Suffix Chains and its Role in Development of the MT System for Azerbaijani

Rauf Fatullayev
National E-Governance Project
Z.Aliyeva str. 33, AZ 1000, Baku, Azerbaijan
Email: fatullayev@gmail.com

Ali Abbasov
National Academy of Science H.-Cavid str., 31, AZ1143, Baku, Azerbaijan
Email: ali@dcacs.ab.az

Abulfat Fatullayev
Institute of Linguistics of National Academy of Science
H.Cavid str., 31, AZ1143, Baku, Azerbaijan
Email: fabo@box.az

***Abstract*— Definition process of the active suffix chains of Azerbaijani (The Azerbaijani language) has been explained in this paper. Morphologically Azerbaijani word-forms consist of the stem and simple or compound suffixes (suffix chains). Because Azerbaijani is an agglutinative language the simple suffixes in this language can form the great number of suffix chains and consequently, it is possible to generate practically "endless" number of word-forms from the same stem. While developing machine translation system from Azerbaijani into non-Turkic languages for the translation of any word-form it is necessary to translate its suffix chain in whole. It is not possible to get the correct translation of a word-form by translating each simple suffix of its suffix chain separately and then by putting these translations together. For these reasons, it is necessary to define the subset of suffix chains frequently used in the texts instead of the set of every possible suffix chains.**

I. Introduction

LANGUAGES of Turkic group (modern Turkish, Azerbaijani, Kazakh, Uzbek, Turkmen, Tatar, Kirghiz and others) are the morphologically rich languages and these languages are characterized by highly productive morphological processes that may produce a very large number of word forms for a given stem. Modeling each word-form as a separate lexical unit leads to a number of problems for the development of formal linguistic technologies as machine translation, speech recognition, text to speech etc. systems.

R esearch into the development of the machine translation (MT) systems for the languages of Turkic group is being carried out in two directions:

1. Development of the MT systems among languages of Turkic group;
2. Development of the MT systems from Turkic languages into languages not belonging to this group and vice versa;

While developing MT systems fro m the languages of Turkic group some problems related to the agglutinative nature of these languages arise. In Turkic languages, every word-form morphologically consists of the stem and the simple or compound suffixes (hereinafter we will call compound suffixes the suffix chains) and as mentioned above, it is possible to generate practically "endless" number of word-forms from the same stem.

Vocabul ary differences are prevalent for the languages of Turkic group. Since the morphological (for example: similar rules of the word formation, existence of the simple suffixes with the similar function) and syntactic structures (for example: the similar formation ways of the noun and verb phrases, the same word order) are very close, it is easier to develop machine translation systems of the first type than those of the second type.

Because the grammatical structures of these languages are very close, it is possible to develop an MT system by creating the Turkic-multilingual dictionary (Table A) of stems and the database of the equivalency (Table B) of simple suffixes [1]-[3].

For example, we presented the translation of the sentence "My little son goes to school" in six Turkic languages.

Mən-im kiçik oğl-um məktəb-ə ged-ir (Azerb.),
Ben-im küçük oğl-um okul-a gid-iyor (Turkish),
Men-inq kiçik ўғл - uм maktab-qa bora-di (Uzbek),
M e n- i n bəlekey bala-m mektep-ke bara jat-ır (Kazakh),
Men-in kiçine uul-um mektep-ke bar-dı (Kyrgyz),
Min-em keçkene ul-um məktəp-qə bar-a (Tatar).

But for the development of the MT system from Turkic languages into (for example) English (analytical language) or Russian (inflectional language) we have essentially different situation.

In these cases we should translate suffix chains in whole (without separating into simple suffixes) (Section 2) and this fact causes some problems. By adding various suffixes to the stem of the same verb, it is possible to create 17947 word-forms in Tatar [4], 11390 in Turkish and 13592 in Uzbek [5]. In the Kazakh language, the number of suffixes that create various word-forms from noun stems is about 500, while most verbs can be used up to 1000 various forms [6]. In Azerbaijani, the number of word-forms formed from the same stem is more than 8000 [7].

For these two reasons (the great number of suffix chains and the necessity to translate suffix chains in whole), it is necessary to define the subset of suffix chains used in texts i nstead of the set of every possible suffix chains.

Despite the great number of suffix chains we should also take into account that the frequency with which all these suffixes and their chains are used is not the same.

If this fact is considered while creating an MT system, then the diff iculties related to the great number of suffix

chains can be avoided. That's to say not all suffix chains, but the suffix chains used in written texts can be determined and included in the database of suffix chains.

TABLE A.
TURKIC MULTILINGUAL DICTIONARY

Azerb.	Turkish	Uzbek	Kazakh	Kyrgyz	Tatar
mən	ben	men	men	men	min
kiçik	küçük	kiçik	bəlekey	kiçine	keçkene
oğl	oğl	ўғл	bala	uul	ul
məktəb	okul	maktab	mektep	mektep	məktəp
ged	gid	bor	bar	bar	bar

TABLE B.
DATABASE OF EQUIVALENCY OF SIMPLE SUFFIXES

im	im	ing	in	in	em
um	um	im	m	um	um
ə	a	qa	ke	ke	qe
ir	iyor	di	ır	dı	a

Suffix chains regularly found in texts, we call *active* suffix chains and our purpose in this paper is to define the subset of active suffix chains of the Azerbaijani.

Hereinafter Azerbaijani is taken as a source and English as a target language and as a translation process, we will mean from Azerbaijani into the language not belonging to the Turkic group.

II. SUFFIX CHAINS AND THEIR TRANSLATIONS

In this section we will show the necessity of taking suffix chains in whole in the translation process.

As mentioned above, compound suffixes that consist of several simple suffixes are called *suffix chains* [8]. We will call the number of simple suffixes that comprise a suffix chain the *length* of this chain. Will also be referred simple suffixes as a suffix chain (whose length is one).

In the grammar of the modern Azerbaijani language, suffixes are divided into two groups—lexical and grammatical suffixes [8]. Lexical and grammatical suffixes form various word-forms from the same word stem by joining word stems both separately and in a certain sequence (for example, it is possible to form the word-forms *ev-də, ev-dəki-lər-in, ev-də-dir-lər, ev-dəki-lər-dən-siniz-mi* and etc. from the stem *ev* , Table 1 *).*

While a word-form is translated into another language, it is necessary to take into account the meaning of each simple suffix that is included in the suffix chain.

At this moment, there is a question: is it possible to get the correct translation of a word-form by translating each simple suffix of its suffix chain separately and then by putting these translations together? It is very important question from the point of view of the development of the MT system from Azerbaijani, but as we will see below , the answer to this question is negative (Section 2).

Let's have a look at the examples shown below (suffixe chains and their translations are underlined in the same way) (Table 2). While the suffix - *də* is translated as *at* and the suffix - *dir* is translated as *he/she/it is* in the first example, in the second example, the suffix - *dir* cannot be translated separately. In this example, the suffix chain - *dir-lər* should be translated as *they are* . In the third example, the suffix chain —*dir-lər* is translated as *they are* and as the plural suffix—*s* which is added to the end of the word. In the fourth example, the rule of translating this chain changes again and the suffix chain - *dir-lər-mi* is translated as *are they* with the plural suffix - *s* added to the end of the word-form. The number of these examples can be increased.

TABLE 1.
DATABASE OF THE ACTIVE WORD-FORMS (FRAGMENT)

Word-form	Word-form
1. ev	*14.* ev-indən
2. ev-dir	*15.* ev-inizə
3. ev-də	*16.* ev-inin
4. ev-dən	*17.* ev-inə
5. ev-i	*18.* ev-lə
6. ev-idir	*19.* ev-lər
7. ev-imdə	*20.* ev-lərdə
8. ev-imin	*21.* ev-lərdən
9. ev-imə	*22.* ev-ləri
10. ev-in	*23.* ev-lərin
11. ev-ində	*24.* ev-lərində
12. ev-lərinə	*25.* ev-lərindən
13. ev-indədir	…

It is necessary to note, that these examples are right not only for English but also for other languages of non-Turkic group (Table C).

TABLE 2.
EXAMPLES TO THE TRANSLATION OF SUFFIX CHAINS INTO ENGLISH

Word-form	Stem of the word-form	Suffix chain	Translation of word-form
1. *evdədir*	*ev* home	*də* - *dir*	he is at home
2. *evdədirlər*	*ev* home	*də* - *dir-lər*	they are at home
3. *tələbədirlər*	*tələbə* student	*dir-lər*	they are student s
4. *tələbədirlərmi*	*tələbə* student	*dir-lər-mi*	a re they student s

TABLE C.
EXAMPLES TO THE TRANSLATION OF SUFFIX CHAINS INTO RUSSIAN

Azerbaijani su ffix	Russian equivalent
-ir, - ır, -ur, -ür (*ged-ir* - he goes)	*– ем , - ём , - ум* (*иð-ем*)
-lər, -lar (*kitab-lar* – books)	*– и , - ы , - а , - я* (*книг-и*)
-ir-lər (*ged-ir-lər* – they go)	*-ут, -ют, -ат, -ят* (*иð-ут*)

As it can be see n from the examples, the compositional translation of each suffix included in the suffix chain can lead to erroneous results. In order to get the right translation, the suffix chain should be translated as a whole, taking into account the meaning of each simple suffix which this chain is composed of.

So, for the development of an MT system, first we would define the set of not all, but only active suffix chains and to develop the translation rules of these chains.

III. ACTIVE SUFFIX CHAINS OF AZERBAIJANI

In the previous section we have show n that in the translation process suffix chains must be taken in whole for correct translation of their meanings. In this section we will define

the set of active suffix chains of the Azerbaijani. Creation of the suffix chains databases is also necessary for the morphological parsing. One of the main purposes of the formal morphological analysis in the Azerbaijani language is to separate the stem of the word-form from its suffix chain, because like all agglutinative languages, grammatical relations among word-forms in the Azerbaijani sentence are determined by the suffix chains and the correct determination of the suffix chain has a serious influence on the correct conduct of the further analysis process.

We should especially point out here that when we say suffixes and suffix chains, we only mean grammatical suffixes and suffix chains formed from them (Number of simple grammatical suffixes of Azerbaijani is about 100 [8]). We are not examining simple lexical suffixes or lexical suffixes in suffix chains, because words formed by lexical suffixes are kept as a separate lexical unit in the dictionary of MT system (this applies both to prefixes – *na, bi, ba, la, a, anti* and to other lexical suffixes - *lı, -li, -lu, -lü, -çı, -çi, -çu -çü, -lıq, -lik, -luq, -lük* etc.).

For example, although the words *balıq* (fish), *balıq-çı* (fisherman) and *balıq-çı-lıq* (fishery) come from the same stem, all three words are kept in dictionary as a separate lexical unit. Because, formally there are not general rules to generate the translation of the word-forms *balıq-çi* (fisherman), *araba-çı* (wagoner) etc. from the translation of the words *balıq* , *araba* (wagon) etc.

Before developing the MT system for the Azerbaijani language, the text corpus on various subjects with more than 300,000 word-forms was created (Later, volume of the text corpus has been increased to more than 12 million word-forms). Number of word-forms found in this corpus was about 39000 (after some processing). These word-forms were put in the database and stems of these word-forms were separated from their suffix chains manually as it presented in the Table 1.

During this process suffix chains were encountered 111,406 times, but the number of various suffix chains was 1,415 (with all variants of the suffix chains with the same meaning). These suffix chains form the basis of the "Database of suffix chains" of the Azerbaijani-English MT system. After grouping these chains (the ones that have the same function, but have different spelling, for example, *acaq, acağ, əcək, əcəy, yacaq, yacağ, yəcək, yəcəy),* the number of suffix chains reached 627.

The length of suffix chains was also calculated during the computer analysis. The arrangement of suffix chains by their length is given in the Table 3 (as we said above, when we talk about the length of suffix chains, we mean the number of simple grammatical suffixes that form this chain). As we can see from these results, very long suffix chains are rare and such chains almost are not used in writing. This can be clearly seen from Table 3. There are no suffix chains longer than five simple suffixes in the texts that were analyzed. The following table shows the frequency with which suffix chains are used by their length.

One of the possible reasons why suffix chains longer than five simple suffixes are not encountered could be that we did not take into account the lexical suffixes. On the other hand, the fact that long suffix chains are not used shows that although the use of such suffix chains is principally possible, no author uses them in writing or if it is necessary, the idea to be expressed by means of a long suffix chain is expressed by a shorter suffix chain (or words) that have the same meaning. For example: the sentence " *Siz bizim dəvət et-di-k-lər-imiz-dən-siniz-mi* " (Are you one of the people who we invited?) is replaced with an equivalent sentence " *Biz Sizi dəvət et-miş-ik-mi* " (Have we invited you?) or another similar equivalent sentence, for example, with the sentence " *Siz dəvət edil-mi-siniz-mi* " (Have you been invited?) A chain that has seven simple suffixes is replaced with a chain that has three simple suffixes.

As it can be seen from Table 3, a chain of five simple suffixes was encountered five times (0.004% of all cases), a chain of four simple suffixes was encountered 223 times (0.200%), a chain of three simple suffixes was encountered 6,895 times (6.189%), a chain of two simple suffixes was encountered 41,331 times (37.099%) and a chain of one suffix was encountered 62,952 times (56.507%).

The distribution of suffix chains that have the same functions without taking into account repetition was as follows (Table 3).

The number of various chains of five simple su ffixes was four, the number of chains of four simple suffixes was 66, the number of chains of three simple suffixes was 248, and the number of chains of two simple suffixes was 257 while the number of chains of one simple suffix was 53.

TABLE 3.
FREQUENCY OF SUFFIX CHAINS

Length of chain	Frequency	Percentage of repeat	Unrepeated chains
5	5	0.004%	4
4	223	0.200%	66
3	6,895	6.189%	248
2	41,330	37.099%	256
1	62,952	56.507%	53
Total	111,405	100.000%	627

Based on these figures, we can say that most of all the su ffix chains used in the Azerbaijani language are consisted of one, two or three simple suffixes.

These suffix chains com pound 99.795% of all most frequent suffix chains. Relative long suffix chains (the ones that have four, five simple suffixes and longer) compound only 0.205% of all chains.

The results that we obtained were analyzed again within the text corpus of 12 million word-forms. Although the volume of the text corpus is increased 40 times, the number of suffixes is increased 1.31 times, while the use of longer suffix chains did not change (that's to say an additional 196 suffix chains were determined and the number of encountered suffixes was 823). After getting this result it is possible to say confidently that active suffix chains of the Azerbaijani language do not exceed 1000.

The fact that the analyzed text corpus has a sufficiently great volume allows us to say that the expansion of the vo lume of the text corpus will not cause a considerable change in relative frequency indicators.

Besides, the types of active suffix chains on the definition of their capability to join the stems belonging to the different parts of speech are also determined. Because, besides well known ambiguity problems (lexical, syntactical etc.), there

are grammatical ambiguity (ambiguity of suffixes) in agglutinative languages else and this information is used in the disambiguation process. 534 chains (≈64.88%) of all chains can join only verb stems (verb chain), 254 chains (≈30.86%) can join non-verb stems (non-verb chain) and 35 chains (≈4.25%) can join both types of stems (dual chain). For the dual chains their frequency of the using as verb or non-verb chains is also defined and this statistics is also used in lexical and grammatical disambiguation process.

So, we have defined the composition of the main information included in the database of the active suffix chains.

The fragmen t of this database is shown in the Table 4.

In the 3rd column of the table 4 are indicated the types of suffix chains. The letter "*V"* written in the third column shows that this suffix chain is a verb chain, while the letter "*N"* - non-verb chain. If none of these letters is written there, the suffix chain is a dual chain.

TABLE 4
DATABASE OF ACTIVE SUFFIX CHAINS (FRAGMENT)

Suffix chain	Other variants of suffix chains	Structure of the chain	Type
am	*əm, yam, yəm*		
da	*də*		*N*
...			
ı r	*ur, ür, yur, yür, ir, yır, yir*		*V*
uram	*ürəm, yuram, yürəm, ıram, irəm, yıram, yirəm*	*ur-am*	*V*
da	*də*		*N*
dadır	*dədir*	*da-dır*	*N*
...			
lar	*lər*		*N*
larda	*lərdə*	*lar-da*	*N*
...			
mış	*miş, muş, müş*		
mışdır	*mişdir, muşdur, müşdür*	*mış-dır*	*V*
...			

In addition, we would like to note that the database of the active suffix chains of the Dilmanc MT system has a more complex structure, but only necessary information used for morphological analysis in examples is presented here.

I n the next section we consider the use of the database of the active suffix chains in the morphological analysis process of the Azerbaijani word-forms.

IV. THE USE OF THE ACTIVE SUFFIX CHAINS DATABASE IN MORPHOLOGICAL ANALYSIS PROCESS

The formation of the grammatical relations among word-forms in a sentence can appreciably differ for different languages. In analytical languages (for example: in English) the grammatical relations among word-forms in a sentence, in most cases, are defined by word order and/or prepositions. In analytical languages, separate words don't have grammatical information and such information can only be acquired in the existence of strict word order. But in agglutinative languages (for example, all the languages of the Turkic group are agglutinative), grammatical relations among word-forms in a sentence are formed by the rich set of suffix chains. For the definition of the grammatical relation among the word-forms at first it is necessary to separate stem and suffix chain of the word-form for the definition of the participation of the word-form in the syntactic structures of the sentence. For this reason morphological analysis algorithms are different for analytical and agglutinative languages.

In agglutinative languages, formal (by computer) morphological analysis can be carried out by creating *a Dictionary of stems* and *a Database of suffix chains*. The dictionary of the Azerbaijani word stems is also developed in frame of the project and in the Table 5 is indicated the simplified version of this dictionary (The dictionary of the Azerbaijani-English MT system has a more complex structure and most of information for the normal functioning of the translation algorithm is not presented here).

Not paying attention to the ambiguity problems, we will schematically describe the work of the morphological analyzer of Azerbaijani.

The morphological analysis algorithm of word-forms in Turkic languages is shown in [7]. This algorithm can be described shortly as follows:

1. The whole word-form is sought in the dictionary of stems (Table 5).
2. If the word-form is not found in the dictionary, its last letter is discarded and the remainder of the word-form – the truncated part is sought in the dictionary again. This process continues until the word-form or its truncated part is found in the dictionary of stems. Discarded part of the word-form is sought in the database of suffix chains (Table 4).
3. If discarded part of the word-form is a suffix chain and this chain can join the stem of this word-form (for example, if the stem is a verb, then the type of the suffix chain should be *V* – a verb chain), then this process stops, otherwise go to the second step;
4. After the stem and suffix chain of the word-form are defined, the word-form is provided with the information included in the databases of stems and suffix chains for its stem and suffix chain.

Example 1. Let's consider the formal morphological analysis process of the word-form *məktəbdədir* (he/she/it is in the school). Starting from the whole word-form all its reminders are sequentially sought in the dictionary of stems (Table 5). Only in the 6th step the stem *məktəb* of the word-form is found in the dictionary. Discarded part - *dədir* is also found in the database of suffix chains. So, process is stopped and we can right

məktəbdədir ⇔ məktəb-dədir.

The word-form and its remainders (according to above mentioned algorithm) with the discarding parts are shown below:

1.	*məktəbdədir*	
2.	*məktəbdədi*	*r,*
3.	*məktəbdəd*	*ir,*
4.	*məktəbdə*	*dir,*
5.	*məktəbd*	*ədir,*
6.	*məktəb*	*dədir*▲

The following examples show how to use the types of suffix chains in the formal morphological analysis process.

Example 2. Let's carry out a formal morphological analysis of the word-form *qorxuram* (I am afraid). According to the morphological analysis algorithm, in the 4th step – the word-form *qorxu* is sought and found in the Table 5. But discarded part of the word-form – *ram* is not suffix chain (Table 4). Therefore the process continues and in the 5th step – the word-form *qorx* (verb stem) is sought and found in the Table 5, discarded part – *uram* of this word-form is sought in

the Table 4 and the process stops because such a suffix chain is found and it is verb chain.

Steps of this process are presented below:

1. *qorxuram*
2. *qorxura* *m,*
3. *qorxur* *am,*
4. *qorxu* *ram,*
5. *qorx* *uram.*

Thus, after this process we get

qorxuram ⇔ *qorx-uram* ▲

Example 3. For the word-form *yazır* (*yaz-ır,* he/she/it writes)

1. *yaz ır*
2. *yazı* *r,*
3. *yaz* *ır.*

in the 2nd step process does not stop, because *r* is not suffix chain (though *yazı* is found in the Table 5). In the next step two variants of the stem *yaz* (verb and noun) are found in the Table 5. Because discarded part – *ır* is a verb chain (Table 4) we chose the verb variant of the stem *yaz*▲

Note that information included in the databases of stems and active suffix chains does not lead to the full solution of the lexical and grammatical ambiguity problem and it is necessary to return to the solution of the ambiguity problem in the next stages of the formal grammatical analysis (syntactic, semantic etc.).

TABLE 5.

DICTIONARY OF DILMANC MT SYSTEM (FRAGMENT)

Stem	Part of speech	English translation
tərcümə et	verb	translate
…		
dilmanc	noun	translator
...		
yaz	verb	write
yaz	noun	spring
yaz ı	noun	record
...		
qur	verb	construct
quru	verb	dry
quru	adverb	dry
quru	noun	land
...		
qorx	verb	play
qorxu	noun	fear
...		
cədvəl	noun	table
...		

V. DILMANC MT SYSTEM

Despite some research works most of languages of Turkic group are still less investigated languages, except modern Turkish [9]-[13].

Researches on the development of Speech and NLP technologies for the Azerbaijani language are being led since 2003 [14]-[16]. Because Azerbaijani is one of less-investigated languages, the most of the necessary works (development of the MT dictionaries, creation of the formal grammar for Azerbaijani, algorithms for the automation of the translation process from/into Azerbaijani, synthesizer and analyzer algorithms of the Azerbaijani sentences, definition of the threephone set for the ASR system etc.) for the development of these technologies are carried out for the first time. The research work presented in this paper is one of such important steps on the creation of the applied linguistic technologies for Azerbaijani (All researches are carried out within the joint projects "Development of the MT system for Azerbaijani", "Development of the Speech Recognition system for Azerbaijani" of the Ministry of ICT of Azerbaijan and UNDP-Azerbaijan).

Dilmanc MT system is a hybrid MT system developed on the basis of RBMT (Rule Based MT) and SBMT (Statistic Based MT) approaches.

Dilmanc MT system can translate for the present on three directions – Azerbaijani-English, English-Azerbaijani and Turkish-Azerbaijani (www.dilmanc.az). For the definition of the factors influencing the translation quality, first the set of test sentences consisting of 1000 sentences is formed. On the results of the test it is possible to say that the system gives good enough - intelligible translations in the most cases (http://www.science.az/cyber/pci2008/1.htm/1-26.doc).

Dilmanc MT system has the following characteristics on each direction (all these items have been developed for the first time):

Azerbaijani-English direction.

1. MT dictionary of Azerbaijani word stems (≈120000 lexical units including word phrases and terms);
2. Database of the active suffix chains (≈1000 active chains);
3. Database of the formalized rules for the decision of the lexical and syntactical ambiguity in Azerbaijani (≈1500 rules);
4. Database of translations of the active suffix chains of Azerbaijani (≈2300 rules);
5. Database of the formal signs of the parts of the sentence in Azerbaijani (≈2000 signs);
6. Formalized rules of the "traditional" grammar of Azerbaijani for the definition of the noun and verb phrases;
7. Formal morphological analysis algorithms of Azerbaijani word-forms;
8. Formal syntactic analysis algorithms of the Azerbaijani sentences;
9. Algorithms for the synthesis of the English sentences.

English-Azerbaijani direction.

1. English-Azerbaijani MT dictionary (≈115000 lexical units including word phrases and terms);
2. Database of the formalized rules for the decision of the lexical and syntactical ambiguity (≈1400 rules);
3. Database of the formalized rules for the synthesis of Azerbaijani suffix chains (≈300 rules);
4. Database of the rules for delimitation of the homogeneous parts in the English sentence (≈90 rules);
5. Database of the rules for delimitation of clauses in the English sentence (≈40 rules);
6. Algorithms for the formal syntactic analysis of the English sentences.
7. Algorithms for the synthesis of the Azerbaijani sentences.

Turkish-Azerbaijani direction.

1. Turkish-Azerbaijani MT dictionary (≈20000 lexical units);
2. Database of the equivalency of Turkish and Azerbaijani suffix chains (≈1000 chains).

It is necessary to note that this list is only a small part of all algorithmic and non-algorithmic means developed in the frame of the Dilmanc MT system.

In addition the formed set of active suffix chains may be used while developing other linguistic technologies as speech and other NLP systems.

Although the analyses are carried out for the Azerbai jani language, there is no doubt that this approach is also applicable for other Turkic languages.

References

[1] Altıntaş K., Çiçekli İ. "A Morphological Analyzer for Crimean Tatar." *In: Proceedings of the 10th Turkish Sy mposium on Artificial Intelligence and Neural Networks,* TAINN, pp. 180-189, North Cyprus.

[2] Cüneyd Tantuğ, Eşref Adalı and Kemal Oflazer, "A MT System from Turkmen to Turkish Employing Finite State and Statistical Methods," in *Proceedings of MT Summit XI,* 2007.

[3] Cüneyd Tantuğ, Eşref Adalı and Kemal Oflazer, "Machine Translation between Turkic Languages," in *Proceedings of ACL 2007–Companion Volume,* Prague, Czech Republic, June 2007.

[4] Iskhakova Kh.F (1968) "Avtomaticheskiy sintez form sushestvitelnogo v tatarskom yazike." *Sovetskaya tyurkologiya,* 2(8): 20-27.

[5] Pines V. Y. (1974) "Nekotorie voprosi avtomaticheskogo perevoda i tyurkskie yaziki." *Sovetskaya tyurkologiya,* 3:100-107.

[6] Bektayev K. (1990) Statistika kazakhskogo teksta. Gilim, Almaati.

[7] Mahmudov M. (2002) Metnlerin formal tehlili sistemi. Elm, Baku.

[8] Abdullayev A., Seyidov Y., Hasanov A. (1972) Müasir Azərbaycan dili (Modern Azerbaijani language). Maarif, Baku.

[9] Cicekli I., Korkmaz T. (1998) "Generation of Simple Turkish Sentences with Systemic-Functional Grammar." In: *Proceedings of the 3rd International Conference on New Methods in Language Processing (NeMLaP-3),* Sydney, Australia, January 1998.

[10] Durgar-El-Kahlout I., Oflazer K. (2006) "Initial Explorations in English to Turkish Statistical Machine Translation." *Workshop on Statistical Machine Translation,* New York, NY, June 2006.

[11] Temizsoy M., Cicekli I. (1998) "An Ontology-Based Approach to Parsing Turkish Sentences." In: *Proceedings of AMTA'98-Conference of the Association for Machine Translation in the Americas,* Lecture Notes in Computer Science 1529, Springer Verlag, October 1998, Langhorne, PA, USA.

[12] Tur G., Hakkani-Tur D., Oflazer K. (2000) " Statistical Modeling of Turkish for Automatic Topic Segmentation ." *Bilkent University, Computer Engineering Technical Report BU-CE-0001,* January 2000.

[13] Vural E., Erdogan H., Oflazer K., Yanikoglu B. (2005) "An Online Handwriting Recognition System For Turkish." In: *Proceedings of SPIE* Vol. 5676 Electronic Imaging 2005, San Jose, January 2005.

[14] Abbasov A., Fatullayev A. (2007) "The use of syntactical and semantic valences of the verb for formal delimitation of verb word phrases." In: *Proceedings of the 3rd La nguage & Technology Conference (L&TC'07).* 5-7 October 2007, Poznan, Poland.

[15] Fatullayev A., Mehtaliyev A., Ahmedov F., Fatullayev R. (2004) "Computer translation system from Azerbaijan language into English." *Proc. of the 4th international conference Internet-Education-Science,* Vinnitsia, 2004, vol. 2, p. 572.

[16] Abbasov A. M., Fatullayev R. A. (2006) "English-Azerbaijani machine translation system on the basis of compressed templates and formal grammatical analysis." *Problems of cybernetics and informatics International Conference (PCI 2006).* Baku – 2006, pp. 42-45.