Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation

Nilufar Abdurakhmonov*

Tashkent State University of the Uzbek Language and literature named after Alisher Navoi

ARTICLE INFO

ABSTRACT

Objective: The paper deals with the significant role of morpheme analysis for modeling of grammatical categories of parts of speech in Uzbek in machine translation. **Methodology:** Identifying types of morphological analysis stages and general paradigms, differences of source and target languages. The verbs in Uzbek have own peculiarities in respect of forms and analytical characteristics. **Results:** There is compound verb, word combination; verbal word combination and their forms play important role for computational morphology. **Conclusion:** The article shows modeling of grammar categories based on forms, bounding of syntactic attitudes and combinations of affixes in forms of verbs.

1. Introduction

One obvious merit of information technologies that cause to develop science and to affect positively optimized infrastructure industry of sphere of knowledge, mainly to connect with broadband communication over the worldwide. Broadly speaking, the systems of each branches of society reformed after established independence in our country. Furthermore, having become progressive changes educational system has been progressed simultaneously since those years. Particularly, spreading in a large scale of computational technologies, namely exposing of opportunity using of Internet system made open the door of the world to face serious issues in science. Most of the directions of sciences crossing with computer have appeared and proceeded. (Agmed, 2015)

As one vivid example is that Computational linguistics turned up in 2000s by affords of professor A.Pulatov where it was established as the first the laboratory at Uzbek philology faculty in National University of Uzbekistan named Mirzo Ulugbek (NUUz). It was basic aim to formulate conceptualization ideology for Uzbek computational linguistics. Those years some topics investigated by the students of master circle of themes modeling, machine translation, automatic orthography, creating of information style of Uzbek. In the direction of machine translation mathematics PhD M.Xakimov (NUUz) has applied many researches on mathematic modeling approaches for multilingual machine translation. The number of application of works done under his supervisor. Some handbooks and course books were created, namely "Computational linguistics" (A.Pulatov, 2011), "The foundations of Computational linguistics" (A.Rakhimov, 2011), "The linguistic basics of machine translation" (N.Abdurakhmonova, 2012) etc. Nevertheless, most implemented works were in theoretical aspect, so there were not any real programs based on full linguistic database. Currently Computational linguistics as a science is being taught in several State universities in Andijan, Namangan, Fergana, Khorazm, Samarkand, Bukhara districts of Uzbekistan. Most of all the center of investigation for CL is Tashkent State University of the Uzbek language and Literature University named after Alisher Navoiy that was established on 13 May 2016 by decree of the first president I.A.Karimov. Despite of fluctuations researches, after the decree for CL it may consider reaching peak of attention to project and recreate the program of BA and MA courses in spite of very short time. Because according to this decree some

essential matters were pointed vividly: "…providing appropriate place of our native language in the Internet world information system, its computational style, scientific-methodic manuals connecting with automatic translator and e-dictionaries, preparing applied recommends and to implement widely results that achieved successfully in practice". We should admit willingly this respectful attitude the Uzbek language, the stream of any kind of research turns only positive side. (Choi, 2015)

The Uzbek language is the language of great Alisher Navoi who founded a rich treasure of the language. As we know, the Uzbek language belongs to Turkic languages and admittedly, it has long-standing history with changeable positions its own destiny by different factors. Its own peculiarities among other languages we can see in every tier of linguistics. For example, as we cite an example saved vowel harmony in the words like üzüm, velâyet in phonetics and kept national words in Turkish. However, there are more loanwords in Uzbek than Turkish. For example, management, budget, test from English, стол, поезд, бухгалтер from Russian, vazir, maktab, maorif from Arabian. On the one hand, lexicology considered very dynamic system deals with social and political situations as well. That is way if we say about Uzbek its grapheme system has also amended several times for centuries. After established independence of Uzbekistan, namely in 1989 the Uzbek language was admitted as state language it had been reforms to improve it constantly. One of them is orthographical rules founded Latin writing settled in 1995. (Hildayanti and Alie, 2016)

One significant issue of computational linguistics in Uzbek to create computational analyzer.

2. Materials and methods

2.1. Abstract of Uzbek grammar

Table 1. Grammar consists two part, namely morphology and syntax.

Parts of speech of Uzbek:			
Dominant elements of a sentence		secondary elements of a sentence	separated groups of the words
Noun	Adverb	Conjunction	Interjections
Verb	Numeral	Auxiliary (Yuklama)	Imitative words
Adjective	Pronoun	Helping words (ko'makchi)	Modal words

The approaching of repsenting grammatical senses, derivation, word formation rules and the forms of formal models in morphology is considered as linguistic procedure. Morphological formal models appeared by usage of word combinations and relations each other in the text. Formal models always exist in the syntagma. Syntagma is semantic-syntactic unit that expresses some unified words as meaningful part of the sentence. Linguistic database involves grammar and dictionary. Generally, parsing implemented by basic three phases during automatically process: (Table 1)

1) Parts of speech
2) Parts of sentence
3) Types of sentence

Uzbek is a morphologically rich language with nouns, adjectives and verbs inflected for case, number and forms of the words. This property requires introducing morphological information inside the MT system to handle the lack of many inflectional forms. It is momentous to create formal grammar of Uzbek for machine translation. (Hosseinzadeh, 2015)

Uzbek has an agglutinative morphology with productive inflectional and derivational suffixes. Because of the suffixes can be added consecutively, one word can convey a lot of information like possessive information, plural/singular, case information, mood etc. Case variation is a widespread linguistic phenomenon. (Kang, 2015)

The literatures that devoted to formal syntax has two major approaches to case assignment can be found. The first approach, which is mainly associated with Noam Chomsky's work, considers case as a syntactic phenomenon that licenses NPs; the second approach, put forward in the work by Alec Marantz, treats case as a post syntactic, purely morphological phenomenon.

There are following models derivation of Uzbek:

W+A=>olma+zor

A+W=>be+foyda

W+W=>tez+yurar

W-W=>ota-ona

W W=>sotib olmoq

W$_{-u/yu}$ W=>Erta-yu kech

Due to lack of grammatical information for natural language processing, it is aim to input descriptive language for the linguistic database.

Modeling of grammatical categories in Uzbek is done in the frame of English for machine translation. English and Uzbek belong to different language family. Therefore, finding unique feature and differences of both languages considered significant matter for morphological analyzer. Let us we see this process as example of the verb in Uzbek.

2.2. THE FORMAL MODEL OF TRADITIONAL MORPHOLOGY IN MACHINE TRANSLAION

It goes without saying that translation process is difficult job due to mental and conceptual matters which exist in different language family, society and cultures of humanity. Linguistical (ambiguity, synonymy) and extralinguistical (psychological) factors affect the quality of the product of translation. Even human translator faces to such hindrance in the process of translation, machine translation comes across these problems as well. As stated above whether related or unrelated language is taken for machine translation, there are some conceptual ideology between languages. Kemal Altıntaş truly estimated as comparing Crimean Tatar and Turkish languages, "the word order and the duties of words in the sentence are most of the time similar. The roots are usually similar, but sometimes they may have different meanings in the two languages". Machine translation among Turkic languages is easier than unrelated languages. (Kim, 2015)

Verbs inflect for number, gender, person and tense, and the two languages share a complex and similar verb structure and inflection system. The two languages share the same verbal forms:

1. The perfective form is used for the past tense in Uzbek

2. The imperfective is used for the future tense in English but is used for a variety of tenses in Uzbek (past, present and future) in coordination with various moods and particles.

3. The imperative

4. The active and passive participles are used for present tense in English and to a lesser extent as a verbal in Uzbek. Verb contains these grammatical categories:

$$
\begin{aligned}
&\text{lexeme} &&->o\'qidim \\
&\text{aspect} &&-> \text{simple} \\
&\text{person} &&->\text{first} \\
&\text{number} &&->\text{singular} \\
&\text{voice} &&->\text{active} \\
&\text{mood} &&->\text{indicative}
\end{aligned}
$$

Figure 1. Verb categories

If one looks at agglutinative languages like Finnish, one finds that morphosyntactic features are encoded systematically by individual morphemes that are arranged in particular linear orders. (Roark and Sproat, 2007)

There are more than 50 affixes give new sense of inflectional verbs, just under 30-word formation affixes; syntactic affixes covered more than 30 forms. Overall qualities verbs in Uzbek consist of more than 6000 words in lexicon. There are about 207 types suffixes (including variation) of parts of speech in Uzbek languages and 130 of them are defined as verbs. In order to add endings to the bases of each words it needs to separate one or another part of speech into paradigms.

In the morphological analysis, stems of words are given in the dictionary with grammatical information and rules. As comparing the stem of "uchmoq" we could see some examples of different models of the following verb structures:

1) Simple verb – uchmoq (fly)
2) Compound verb – uchib ketmoq (fly away)
3) Collocation – samalyot uchirmoq (fly the plane)
4) Verbal word combination – varrak uchirib bermoq (fly the kite to smb.)
5) Modal word combination – uchirish kerak (must fly)
6) Idioms – kapalagi uchib ketmoq (be afraid)

there are two types of analytical forms of the verbs in Turkic languages:

- analytical forms of conditions;
- analytical forms of modality;
- analytical forms of other mood.

Apart from these, morphological analyzer should parse correctly each segments in the text. Otherwise, some homonymic problems surface in the translation of the units in the text. For instance verbal word combination qo'yib berdi is used in many functions as homonym in the context like the following examples:

U hujjatni stolga qo'yib berdi-> He gave document as putting on the table.

U bolani hovlida o'ynab olishiga qo'yib berdi-> He let the boy play in the yard.

Direktor ko'rsatilgan hujjatlarga darhol imzo qo'yib berdi-> The director signed abruptly brought documents.

U bolalar o'ynab olsin deb, sho'x ashula qo'yib berdi-> He played music so that to dance the children.(Leung, 2016)

Database and semantics of verbal word combinations are investigated very little even in Uzbek. Considering all of them main verb, there are not any pure verbal helping verbs that they are used independently. They are about 30 types of the verbs but they give different meanings to the notional verbs. They

are: ber (ver), bil, bit (bitir), bor, boshla, boq, bo'l, et, yoz, yot, kel, ket, ko'r, ol, sol, tashla, tur, tush, chiq, yubor, yur, o'l, o't, o'tir, qara, qol, qo'y. There are two types model of verbal combinations:

- (MV+b/ib) HV
- (MV+a/y) HV

25 helping verbs combined with main verbs with affix -(i)b:

	berdi	bo'ldi	ko'rdi	(yoqib) tushdi	o'tdi
O'qib	Bildi	(anglab) etdi	Oldi	chiqdi	o'tirdi
	bitirdi	yotdi	(aytib) soldi	yubordi	Qaradi
	bordi	keldi	tashladi	yurdi	Qoldi
	boqdi	ketdi	Turdi	o'ldi	qo'ydi

11 helping verbs joined with affixes -a/y :

1) ber/ver (yoza berdi)   5) yoz (yiqila yozdi)   9) sol (kela solib)
2) bil (topa bildi)   6) ket (gapira ketdi)   10) tur (yoza tur)
3) bor (o'zgara bordi)   7) ko'r (ayta ko'rma)   11) qol (ayta qol)
4) boshla (yoza boshladi)   8) ol (unuta oldi)

16 helping verbs cannot join with main verbs via affix -a/y, they only link each other by means of affixes b/ib: bit (bitir), boq, bo'l, et, yot, kel, tashla, tush, chiq, yubor, yur, o'l, o't, o'tir, qara, qo'y.

9 affixes can join both of the forms of affixes (b/ib and a/y): ber (aytib ber, ayta ber), bil (yozib biladi (in dialect), yoza biladi), bor (o'qib bordi, o'qiy bordi), ket (isib ketdi, gapira ketdi), ko'r (aytib ko'r, ayta ko'rma), ol (yozib ol, yoza ol), sol (to'kib soldi, kela solib), tur (o'qib tur, yoza tur), qol (tugab qoldi, ayta qol). However these verbs can join both of the affixes, but only they have different meanings: yozib oldi (take a note)– yoza oldi (could write), aytib ko'r (try to tell) – ayta ko'rma (don't tell any more). Sometime they have the same meanings: og'irlashib bordi – og'irlasha bordi.

The meanings of verbal word combinations: (Table 2)

Table 2. Meanings of verbal word combination

form	meaning	Example
-(i)b ber	Direct activity to another person	qo'yib ber, bilib ber
-a/y ber	Continuity	ko'chaver, o'ylayver
-a/y ol/bil	Capability	tuzata oldi, foydalana bil
-(i)b bit/bitir	Perfect aspect	yonib bitgan, ekib bitir
-i)b, -a/y bor/kel	Continuity	unutib bordi, o'zgartira bor
-a/y boshla	Beginning	yura boshla, o'qiy boshla
-(i)b boq/ko'r/qara	Proof	o'qib boq, o'ylab qara, yozib ko'r
-(i)b bo'l	Perfect activity	yozib bo'l, yuvib bo'l, qazib bo'l
-(i)b et	Perfect activity	anglab etdi, pishib etmagan
-a/y yoz	Close to activity	qula yozdi, yorila yozdi
-(i)b yot/tur/o'tir/yur	Continuity	o'qib yot, aytib tur, yozib o'tir
-(i)b ket	Beginning and Continuity	tarqab ketdi, isib ketdi, g'ovlab ketdi, maqtab ketdi
-a/y ket	Beginning and Continuity	o'qiy ketdi, gapira ketdi
-(i)b sol	Activity perfect	aytib sol, yuragini to'kib soldi
-a sol	Activity in turn	ola solib ot, kela solib boshla, tura solib tashlan
-(i)b tashla/yubor	Full and fast appeared activity	o'qib tashla, to'qib tashla, haydab yubor
-(i)b chiq	Finish	o'qib chiq, so'rashib chiq, aylanib chiq
-(i)b tush	Completely	ag'darilib tushdi, yiqilib tushdi, yoqib tushdi, yarashib tushdi
-(i)b o'l	Continuity and Repeated	mudrab o'lyapman, surishtirib o'ldi, sog'inib o'ldi, chanqab o'ldi
-(i)b o't	Perfect	bo'lib o'tdi, gapirib o't
-(i)b qol	Unexpected and continuity	tikilib qol, serrayib qoldi; kelib qoldi, so'rab qol; jonlanib qoldi
-a/y qol	Permission, agree, wish	bora qol, yura qol, keta qol
-(i)b qo'y	Continuity and once time activity	suyab qo'y, ilib qo'y; o'ylantirib qo'y, shoshirib qo'y; qarab qo'y, yo'talib qo'y

Helping verbs are sometimes written as shortened forms or compound verbs: aytib yubor – aytvor, bora ber – boraver, unuta olmadi – unutolmadi.

In contains of such kind of verbs there are several verbs: aytib <u>berib tura qol</u>, aytib <u>bera olmay qoldi</u>. Every time the first part of the verbs is considered main verb: Aytib (main verb) ber (helping verb), aytib (main verb) berib tur (helping verb). (Kim, 2015)

Verbal word combination looks like to compound verbs, but only the first component of the verb give main meaning others help to this main verb, in compound verb save both of the components save independent meanings. We see three the same model:

1. (MV+PP) HV=> oshib tushmoq (Compound verb)
2. (MV+PP) HV=> oshib ketmoq (Verbal word combination)
3. (MV+PP) HV=> to'pni oshirib uzatmoq (Collocation)

1) if taken off helping verb from notional verbs, so combination may have a few change in the meaning: yozib bordi (write on) – yozdi (write), isib ketdi (warm up) – isidi (get warm). Nevertheless, taken off helping verb from compound verb it will be entirely changing of sense word, because helping verb participate to derivate a new word: sotib ol (buy) – sot (sell); ishlab chiqardi (produce) – ishladi (work);

2) there are more than two independent units of collocation in the text: Quvonib (modifier) so'zladi (predicate) – he talked joyfully; in verbal word combination there is only one predicate: O'ylab qoldi (predicate).

The affixes of voice and negative form could be added both parts of verbal word combination: to'xtatib qo'y<u>ishdi</u>; aytib qo'y<u>ma</u>, ayt<u>may</u> tur, ayt<u>may</u> tur<u>ma</u>; the affixes of tense, mood, person are joined to helping verbs. Apart from these syntactic forms are existed simultaneously in both parts of the verb: tamomla<u>shdi</u>-qo'y<u>ishdi</u>.

It is customary in discussions of morphology to talk about inflectional versions derivational morphology in Uzbek, in terms of the types of features each of these encodes.

It is important issue that modeling of grammatical forms in machine translation:

1) Modelling of verbal word combinations;
2) Finding adequate sense of verbal word combinations in English.

These are general models for verbal word combinations:

- MV HV->o'qib berdi
- MV+HV->berolmadi<=> bera olmadi
- [MV]-[HV]->yozdi-qo'ydi
    1.1. Here is some peculiarities for modelling:

MV-main verb (MV – such main verbs keep own notional meanings)

HV_1-helping verb (HV_1 – such verbs are added after -b/-ib affixes)

HV_2-helping verb (HV_1 – such verbs are added after -a/-y affixes)

- MV+HV_1=>aytib berdi
- MV+HV_2=>so'zlay oldi

Verbal word combinations are similar to phrasal verbs in English (look up, look forward) that some of the preposition or adverbs give additional meanings to notional verbs (Table 3)

Table 3. some of the preposition or adverbs give additional meanings to notional verbs

o'qib bo'l-	Mushtariy kitobni o'qib bo'ldi.	Mushtary has finished the book.
o'qib chiq-	Mushtariy kitobni o'qib chiqdi.	Mushtary had read the book through.
o'qib tur-	Mushtariy kitobni o'qib turdi.	Mushtary used to read the book.
o'qib yubor-	Mushtariy kitobni o'qib yubordi.	Mushtary read the book suddenly.
o'qib tashla-	Mushtariy kitobni o'qib tashladi.	Mushtary easily has read the book.
o'qib ol-	Mushtariy kitobni (qayta) o'qib oldi.	Mushtary read the book one more.
o'qib ko'r-	Mushtariy kitobni o'qib ko'rdi.	Mushtary tried to read the book.
o'qib qo'y-	Mushtariy kitobni (o'zi uchun) o'qib qo'ydi.	Mushtary read the book for herself.
o'qib ber-	Mushtariy kitobni (ukasiga) o'qib berdi.	Mushtary read the book her brother.
o'qib ket-	Mushtariy kitobni (to'xtamay) o'qib ketdi.	Mushtary read the book without no pause.

Nevertheless, similarity in both languages, there is one significant angle of phrasal verbs changed definitely the meanings unlike verbal phrasal verbs in Uzbek. In addition, it cannot compare two categories as the morphological unit owing to own specificity of languages and the models of verbal word combination according to coming in what and how position are different:

- MV+HV=> ko'rib qoldi
- MV+$HV_{(1)}$ + $HV_{(2)}$ =>ko'rsatib bera oldi
- MV+$HV_{(1)}$ + $HV_{(2)}$ +$HV_{(3)}$ =>berib qo'ya qoldi
- MV+$HV_{(1)}$ + $HV_{(2)}$ +$HV_{(3)}$ +$HV_{(4)}$ =>aytib berib qo'ya qoldi
- MV+$HV_{(1)}$ + $HV_{(1)}$ =>o'qib tura tur
- $MV_{(1)}$ +$HV_{(1)}$ =>tura tur
- $MV_{(1)}$ +$HV_{(1)}$ +$HV_{(2)}$ =>turib tura qolgin

As we pointed above, some models [verb + verb] include the same roots may come several times and they give separately meanings in the text.

Moreover, they also look like compound verbs with external form in accordance with structure:
-    Oshib tushmoq->Compound verb (climb over)
-    Oshib ketmoq->Verbal word combinations (rise up)

On the other hand, helping verbs come as the component in compound verbs and idioms as well:
-    Nonushta qilmoq (have a breakfast) =>nonushta qilib berdi(Compound verb);
-    Mashq qilmoq (do exercise) =>mashq qilib turdi (Compound verb);
-    Kapalagi uchib ketmoq (to be afraid) => kapalagini uchirib yubordi (Idiom);

Morphologic analyzer should identify each units properly according to their contextual meaning. In Uzbek verb has following predicative forms:

Negative Form-N {-ma| -mas| -may}

Conditional form-CF {-sa}

Tense-T {-a|-y|-yap|-moqda|-yotir…}

Predicative form-PF {-man|-san|-dir…}

Voice-VF {-t|-tir|-giz, |-kiz…}

Non-finite form-NF {-gan|-kan|-qan|-b|-ib…}

Person-P {-im|-ing|-k|-ngiz|-lar| -man| -san|-k|-ngiz}

Purpose-PS {-moqchi}

COMBINATION EXAMPLE
-    MV+ N+PF=> O'qimayman
-    MV+ N+ T+PF=> O'qimadim
-    MV+ CF +PF=> O'qisam
-    MV+ N +CF +PF=>O'qimasam
-    MV+ T+PF=>O'qiyapman
-    MV+VF=>O'qittir
-    MV+NF=>O'qigan
-    MV+PS+PF=>o'qimoqchiman

Affixes could be added in both of components of verbal word combination:
-    (MV+N+NF) +HV=> ko'rmay qoldi
-    (MV+ NF) + (HV+ N) => ko'rib qolmadi
-    (MV+N+NF) + (HV+ N) => ko'rmay qolmadi
-    (MV+V+NF) +HV =>ko'rsattirib qo'ydi
-    (MV+NF) + (HV+ PF) =>ko'rib borsam
-    (MV+NF) + (HV+ N+PF) =>ko'rib bormasam
-    (MV+NF) + (HV+ VF+N+T) =>yeb ko'rgizmadi

Apart from above mentioned all forms of verbal word combination there are some structures of verbs edi, ekan, emish and bo`lmoq, hisoblanmoq, sanalmoq, deyilmoq. These linking verbs (these verbs create predicative forms) formulate predicative forms of the verbs they have following models:

a)   N|Adj.|Num.|Pron.|Adv.+bo`lmoq=> Agar soat o'n bo'lsa, sizga qo'ng'iroq qilaman (If it is 10 o'clock, I shall call you);

b)   yo`q/bor/oz/ko`p/zarur/lozim/kerak+bo`lmoq=> O'ylagan orzularim bir pasta yo'q bo'ldi (My thinking dreams disappeared abruptly).

d)   Infinitive+kerak/lozim/shart/darkor: xona tozalanishi kerak=> Xona tozalanishi kerak (The room needs cleaning. - The room needs to be cleaned).

2.3. DATABASE OF PHRASAL VERBS AS ANALITIC MODELS IN ENGLISH INTO UZBEK TRANSLATION

For building machine translation system from English into Uzbek there should be amount the measure of computer-assisted dictionary that saved in database. Both English and Uzbek have very huge database consisting all linguistic layers and they are very dissimilar. One verbal category in English is phrasal verb. Somehow, phrasal verbs in English like verbal word combination in Uzbek, however it has own peculiarities as well. It is a challenge for structural components of sentence. "Phrasal verbs are considered to be a very important and frequently occurring feature of the English language. Firstly, they are so common in every - day conversation, and non - native speakers who wish to sound natural when speaking this language need to learn their grammar in order to know how to produce them correctly. Secondly, the habit of inventing phrasal verbs has been the source of great enrichment of the language. By means of phrasal verbs it is described the greatest variety of human actions and relations". Therefore, verb-particle constructions in English are very complex to analyze and describe coherently in synchronic terms.

Database design is considered to be of the stage of cycle information system and very crucial task is normalization of each unit in the process of database design. We have gathered more than 12 thousand phrasal verbs and their 3 thousand separate notional verbs. Nevertheless, every unit placed in detached cell in database and overall qualities of them is more than 80 thousand phrasal verbs and notional verbs. This proposed translation the direction of English-Uzbek-English.

ID	English	Transcription	(Pro)Noun	Sighns	1 (Pro)noun	Uzbek	Translation	1 Synonims	2 Synonims
ID	Inglizcha	transkripsiyasi	noun, pronoun	belgilar	noun, pronoun	O'zbekcha	sinonim	sinonim	sinonim
1	abandon	[ə'bændən]				bosh tortmoq	rad etmoq		
2	abandon	[ə'bændən]				tashlab ketmoq	tark etmoq		
3	abandon	[ə'bændən]				qoldirmoq	berib yubormoq		
4	abandon		oneself	to	smth.	berilib ketmoq	o'zini bag'ishlam		
5	abandon		smb. or smth.	to	smb. or smth.	tashlab ketmoq	tark etmoq	qoldirmoq	
6	abbreviate	[ə'bri:vıeıt]				qisqartirmoq	kamaytirmoq	kichraytirmoq	
7	abbreviate		smth.	to	smth.	+gacha qisqartirmoq	kamaytirmoq		
8	abbreviate		smth.	as	smth.	+gacha qisqartirmoq	kamaytirmoq		
9	abduct	[ab'dʌkt]				o'g'irlamoq	olib qochmoq		
10	abduct smb. fr					(odam) o'g'irlamoq	olib qochmoq		
11	abet	[ə'bet]				muhtojlikda yordam bermoq	qo'llab-quvvatla		
12	abet		smb.	in	smth.	(yomon ishga, jinoyatga) undam			
13	abide	[ə'baıd]				kutmoq	intizor bo'lmoq		

Figure 2. Translation the direction of English-Uzbek-English.

In data table with the transcription of main verb will give opportunity aids use in e-dictionaries. In addition, nearly all meanings of main verb with their synonyms in the same line are included into database. It helps users to search for all synonyms of verbs not only the main meanings but also secondary meanings of verbs as well of that of phrasal verbs. Some separable or inseparable phrasal verbs are taken into account too. This database formed according to following models of phrasal verbs (V-verb; P-particles (preposition or adverb); ↴ -possible or impossible; smth.-something, smb.-somebody)):

- V + oneself +P+smth. => align oneself with smb. or smth.
- V+ oneself +P=> arch (oneself) over
- V + oneself +P+smb. => attach oneself to smb
- V + P+ smb. or smth. + P smth. =>arrange with smb. about smth.
- V+ smb. or smth. + P+ smb. or smth. => associate smb. or smth. with smb. or smth.
- V+ smth. +P+ smth. => balance smth. against smth.
- V+ smth. +P=> bail smth. out
- V+ smb. +P+ smth. => astound smb. with smth.
- V+smb. +P+ smb. or smth. => bias smb. against smb. or smth.
- V+smb. +P=> beat smb. up
- V +P+smb. => bet with smb.
- V +P+smb. or smth. => attend to smb. or smth.
- V +P+ smth. +P+ smth. =>
- V +P+ smb. +P+ smb.or smth.=> book smb. through (to some place)
- V+P+P=> be in for
- V+P+P+it=> be in for it
- V+P+P+smth=> be off for smth.
- V+P+P+ smb.or smth.=> bound up with smb. or smth.
- V+P+ P+ smth=>bear up (against smth.)

And we also included some symbols to clarify each other's. They are followings:

- P1, P2, …, PN –N meanings of main verbs;
- N1, N2, …, Nk –k forms of phrasal verbs;
- B1, B2, …, Bl –l meanings of phrasal verbs;
- P1_, P2_, …, PN_ - synonyms of main verbs;
- B1_, B2_, …, Bl_ – synonyms of meanings of phrasal verbs;
- P – this means it doesn't belong to any main verb.

After having investigation of phrasal verbs we may conclude that phrasal verbs corresponds with simple verbs, compound verbs, word combination, verbal word combination and idioms in Uzbek.

According to Yorick Wilks "While we agree that is unlikely that the information in machine-readable dictionaries is sufficient for this grand database of facts that will support NLP as a whole, we are optimistic about making use of the information they do provide to support the creation of lexical entries for specific natural language processing systems" (Wilks, 2009)

2.4. VERBS IN UZBEK FOR MORPHOLOGICAL ANALYSIS SYSTEM

Automatic morphological analysis dates back to the earliest work in computational linguistics on Machine Translation during the 1950s (Andron, 1962; Woyna, 1962; Bernard-Georges et al., 1962; Boussard and Berthaud, 1965; Vauquois, 1965; Schveiger and Mathe, 1965; Matthews, 1966; Brand et al., 1969; Hutchins, 2001). There have been many applications over the years including the Porter stemmer (Porter, 1980) heavily used in information

retrieval applications (Dolby et al., 1965; Attar et al., 1978; Choueka, 1983; Büttel et al., 1986; Meya-Lloport, 1987; Choueka, 1990; Koskenniemi, 1984), spelling correction (McIlroy, 1982; Hankamer, 1986), text input systems (Becker, 1984; Abe et al., 1986), and morphological analysis for text-to-speech synthesis (Allen et al., 1987; Church, 1986; Coker et al., 1990). Many of these earlier applications used quite ad hoc approaches including hard-coding much of the linguistic information into the system. For example, in the system reported in Coker et al. (1990), a lot of the morphological analysis is mediated by tables coded as C header files and spelling-change rules written as C functions (Rorak and Sproat, 2007)

It is necessary Word manager in machine translation. Word Manager (WM) is a system for morphological dictionaries. A morphological dictionary is a database in which lexical knowledge is organized according to morphological rules.

From our point of view, morphological analysis must not be restricted not only for parsing categories, mainly in machine translation. As for Radolfo Delmonte contributes linguistic categories in Italian following types (Rodolfe, 2008)

➢ grammatical categories - derived from a categorization of reality into entities - nouns -, events - verbs and nominals -, attributes - adjectives, adverbials, and nouns;

➢ semantic categories, like negation, quantifiers;

➢ discourse level categories, like deictics, definiteness, conjunctions for coordination and subordination at propositional level;

➢ syntactic categories - encoding the arity of predicate-argument structures as they are interpreted in situations;

➢ aspectual categories - encoding the internal temporal structure of events (as
expressed both by verbal and deverbal nominals);

➢ semantic conceptual categories - classifying types of events in relation to the
(un)reality they encode;

➢ selectional restrictions - encoding the typicality of event participants in inherent semantic features as they are represented in an ontology or connected encyclopedic database of entities and their semantic interrelationships;

➢ grammatical constraints - encoding socalled syntactic and anaphoric binding constraints on arguments of predicate and dependent predicates only for propositional arguments, though.

As we see, grammatical categories are the major part of morphological analysis. According to some works, automatic analysis has following types

- Analyzing stem in lexicon
- Analyzing word forms in lexicon
- Analyzing through logical approach
- Analyzing via the tables without lexicon

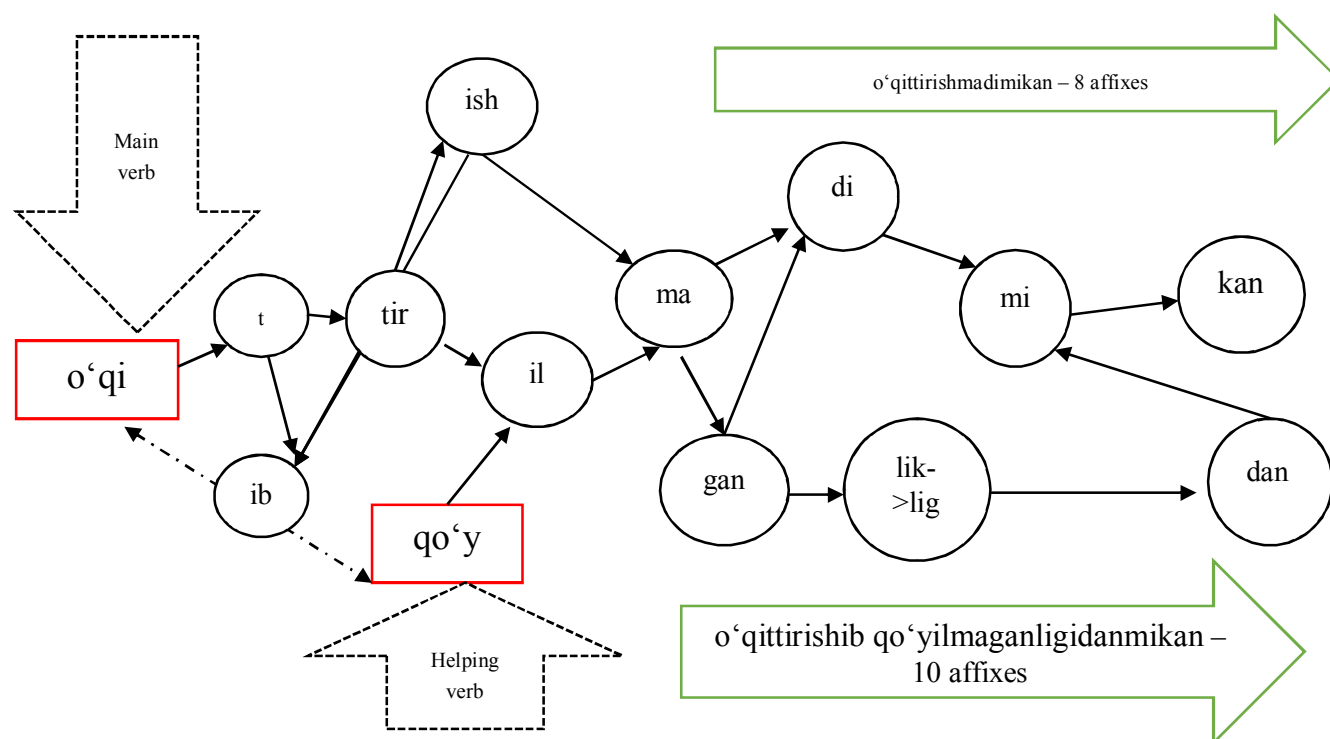Which of type is appropriate for Uzbek, let us to next algorithmic system of affixes of verb: (Figure 3)

Figure 3. Algorithmic system of affixes of verb

Here some verbs like root <koʻn> but the give partly definitions: koʻnmoq<=>koʻnikmoq

1. V->koʻn (agree) -> U mening shartimga koʻndi (He agreed to my condition)
2. stem +ik=Verb ->koʻnikmoq (used to)– Men shahar hayotiga koʻnikdim (I am used to living in urban lifestyle)

Most of all it should be directed verbs in morphological analysis morphemes which including in lemma. "Lexicon entries are tokenized using a simple left-to-right longest match tokenizer algorithm. The entry is tokenized by going through the entry string, position by position, and looking up the longest symbols available using a very simple greedy tokenizer. If the tokenizer is incremental, it memorizes new tokens as it parses the input assuming that multicharacter tokens have been declared in advance. An alternative, but less efficient, strategy is to determine all the tokens in a separate pass in order to compose the entry string with a tokenizer-transducer implementing a greedy left-to-right matching or some other strategy to achieve the desired partitioning's. Let us to cite one more example to analyze the verb in the text: Men hali ham <u>tuzalganim yo'q</u>. –If automaton announce underlined word as predicate, the core problem is how it will be take apart morphemes correctly so that given proper meaning into another target language.
TUZALMOQ: 1) recover, be cured; 2) lay the cloth; 3) be repaired; 4) be corrected
 As we see, there are a few meaning of the word "tuzalmoq", so we should identify the appropriate sense of lexeme in accordance with contextual position.
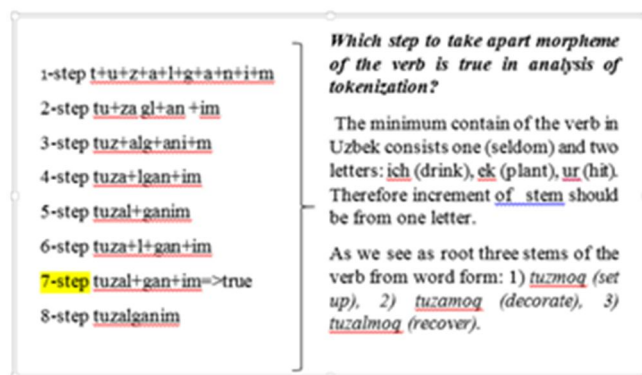

Figure 4. appropriate sense of lexeme in accordance with contextual position.

Word manager seek the list of affixes in database. Hence it will be obvious what affixes are correct form that may be added the verb. It has like database list: (Table 4)

Table 4. Database list

ID	Affixes	Function	Abbreviation
1.	ga	Case	C
2.	gan	Tense (Past)	T_{past}
3.	gan	Participle (Past)	PP
4.	l	Voice (Active)	$V_{act.}$
5.	l	Voice (Passive)	$V_{pass.}$
6.	a	Tense (Present)	$T_{pres.}$
7.	m	Person (first)	P_1
8.	i	Possessive pronoun (3-person)	$PossP_3$
9.	im	Possessive pronoun (1-person)	$PossP_1$

After checking affixes, it is tested the module of word formation and combination affixes.
- W_{stem}+V+PP+PossP=>tuza+l+gan+im=> Passive voice
- W_{stem} +PP+PossP=>tuzal+gan+im=> Active voice

Here it may be appeared problem whether which of them is true. Furthermore, taking into consideration analysis of these data, semantic analysis is required after this process. Assume for the sake of argument that affixes give different sense with various functional positions. Negative forms of the verb are also considered one important paradigm in Uzbek. Because variation forms cause to change meaning as well. Following models are for indicated as negative forms in different collocations:

I.   MV+ma=>o'qimadi (He did not read)
II.   (MV+may) HV=> O'qimay qo'ydi (He is not used to read)
III. MV (HV+ma) =>O'qib qo'ymadi (He did not read any more)
IV.   (MV+may) (HV+ma) =>O'qimay qo'ymadi – positive meaning (Surely he read (in the past))
V.   (MV+PP) +emas=>O'qigan emas (He has never read)

VI.   (MV+PP+Poss) yo'q=>O'qigani yo'q (He did not read)
VII.   Na MV {CV, MV, VC} =>na habar oldi | na o'qidi | na berib ketdi (Neither he read)
VIII.   (MV+mas) edi=>O'qimas edi (He did not use to read)
IX.   (MV+ma+gan) +ekan+P=>O'qimagan ekanman (I did not read)
X.   Na (MV+ma) =>Na o'qimadi (He did not read)

3. Discussion and results

Composition of regular relations is the single most general computational operation that can handle the formal devices found in natural language morphology.

Problem is one that there is not Uzbek language among other MT systems (Solver.uz, Google.translator etc.). Moreover, none of them can translate properly yet from Uzbek into English and vice versa. Just taken as example we may compare how correctness realized among parts of speech in the text: (Figure 5)
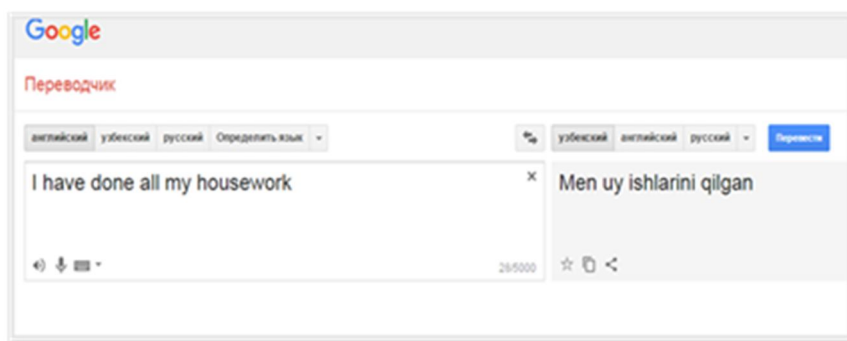

Figure 5. how correctness realized among parts of speech in the text

Estimated one that semantic pole of lexicon and contextual meaning of the word forms in machine translation system is very indispensable for linguistic database. Owing to lack of words and affixes combinations and not given appropriateness, grammatical categories in source and target languages cause only e-dictionaries but not machine translation. In the text, translation in Uzbek should have been: Men barcha ishlarimni qilib bo'ldim. Let us we analyze similar text in context free grammar: (Figure 6)
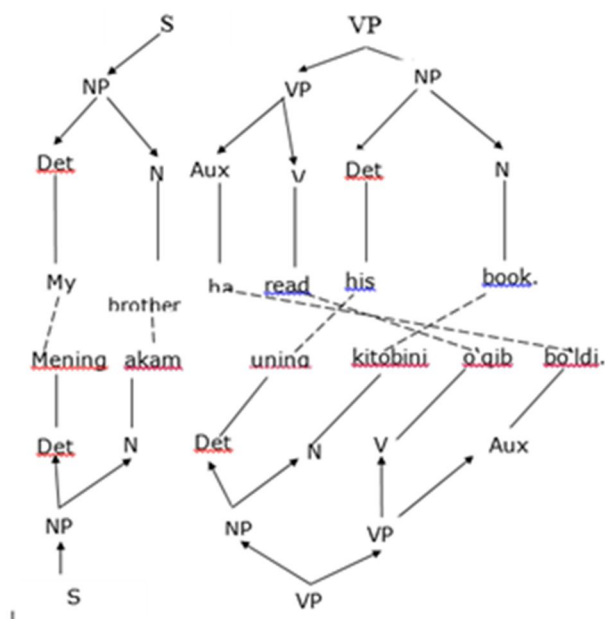

Figure 6. Text in context free grammar

As visualized the model sentence, it should be clear how parts of sentence related each other in turn both languages. Particularly there are not some categories in each language shown in the chart. Corresponding parts of speech and semantic poles of words are should be taken into account in machine translation as well.

Moreover, we could see in the table 5 how do parts of speech correspond between languages:

Table 5. How do parts of speech correspond between languages

English –Uzbek – English translation					
	Kecha	ba'zi	Talabalar	darsga	kelmadilar
Yesterday	■				
some		■			
of			■		
the			■		
students			■		
did					■
not					■
come					■
to				■	
the				■	
lesson				■	

Mainly an amount of researches based on statistical MT. The fact that probabilistic and mathematical approaches to MT are useful, but somehow the essence of natural language should be not escaped like problems. Particularly English and Uzbek languages are separable particularities owing to linking of different language families. Machine-readable dictionaries are available now to adopt to MT, but then again to create linguistically database should be improved as we stated above. It is very important to build database simultaneously English language with Uzbek. Otherwise, it translates erroneously from source language into target language. On the other hand, it is very core problem to create semantic database with conformable to contextual meaning of the word combination. Consequently, we offer to provide to give database of appropriate valences word combinations in ample lexicon:

read somebody like a book – biror kishini juda yaxshi tushunmoq

book club – kitobxonlar to'garagi

speak by the book – aniq ma'lumotga tayanib gapirmoq

If our investigation based on phraseological principles with blended approach of rule-based translation, machine translation will be available to give proper meaning to conceptualizing whole sense of the discourse. We have another one problem that homonymy of word combination and sentences. For example, It is a piece of cake – 1) Bu tort bo'lagi; 2) Bu juda oson. Hence, what should we do to solve such problems? In perspectives, it will be required parallel corpora in this sphere. However, to begin researches making corpus we use the structure of sentence.

Having completed the morphological analysis finding the appropriate each models between languages, the text is formulated over the analyses so that to save the context information. In many cases, the order in which the lexemes appear is significant and the meaning of the text is directly dependent on that order. When the words are ambiguous, namely the units compounded several parts like collocations, idioms or phrasal verbs, all possible combinations with the words are generated simultaneously.

4. Conclusion

In general, linguistic models and semantic relations of each unit in machine translation play important role to build database. Due to globalization process, everything is getting to change into apace; additionally, there is not any hindrance to unify cultural and social attitudes among people. So understanding other foreign language has become a very crucial thing that we cannot ignore it. Today the result of machine translation which was appeared last in the mid of century impacted expand researches in this sphere one is the direction of computational linguistics and state of the art information technologies give opportunity to use them in any kind of branches of public. As far as we know that, a good machine translation is one of the tools of the trade for translation procedure. Using only grammatical models cannot solve all problems in machine translation system. It goes without saying that subsumption of entries into dictionaries according to grammatical information, helps to seek in the linguistic database. However, if it keeps going not attentively full information of language, artificial intelligence is still stay only translator toolkit during translation process.

REFERENCES

Ahmed, S. 2015. Child with Intellectual Disability and Parental Response: A Comparative Study of Nuclear and Extended Family Structures in Khyber Pakhtunkhwa, Pakistan. Humanities & Social Sciences Reviews, 3(1), 30-33. Retrieved from https://giapjournals.com/index.php/hssr/article/view/241

Choi, S. 2015. Re-examining the Dimensionality of Leisure Motivation and Leisure Satisfaction in a Multicultural Context: Evidence from Macau. Humanities & Social Sciences Reviews, 3(1), 06-10. Retrieved from https://giapjournals.com/index.php/hssr/article/view/238

Hildayanti, S., & Alie, J. 2016. FACTORS INFLUENCED PADDY FARMERS TO USE OR NOT USE ORGANIC FERTILIZERS IN SOUTH SUMATERA, INDONESIA. Humanities & Social Sciences Reviews, 4(1), 53-58. https://doi.org/10.18510/hssr.2016.417

Hosseinzadeh, V. 2015. Salafism and its Socio-Intellectual Foundations. Humanities & Social Sciences Reviews, 3(1), 34-37. Retrieved from https://giapjournals.com/index.php/hssr/article/view/242

Kang, H. 2015. The Effects of a Non-traditional Monetary Policy on Housing Prices in Taiwan. Humanities & Social Sciences Reviews, 3(1), 38-41. Retrieved from https://giapjournals.com/index.php/hssr/article/view/243

Kim, B. 2015. Colonial Modernity and Tourism in East Asia. Humanities & Social Sciences Reviews, 3(1), 16-29. Retrieved from https://giapjournals.com/index.php/hssr/article/view/240

Kim, C. 2015. The Birth of a Political Activist: A Comparison of Two South Korean Films. Humanities & Social Sciences Reviews, 3(1), 01-05. Retrieved from https://giapjournals.com/index.php/hssr/article/view/237

Leung, C., & Chan, W. (2016). SOCIOLINGUISTIC PHENOMENON OF CODE MIXING IN HONG KONG: FROM A PERSPECTIVE OF MARKETING COMMUNICATIONS. Humanities & Social Sciences Reviews, 4(1), 20-26. https://doi.org/10.18510/hssr.2016.413

Roark, B. & Sproat, R. 2007. Computational Approaches to Morphology and Syntax. Oxford University Press Inc., New York, 63.

Roark, B. & Sproat, R. 2007. Computational Approaches to Morphology and Syntax, OXFORD, 100.

Rodolfe, D. 2008. Computational Linguistic Text Processing: Lexicon, Grammar, Parsing and Anaphora Resolution. Nova Science Publishers, Inc. New York,4-5.

Wilks, Y. 2009. Machine translation. Its scope and limits. Spring science+Business Media LLC. UK, 139.

How to Cite this Article:

Abdurakhmonov N., Modeling Analytic Forms of Verb in Uzbek as Stage of Morphological Analysis in Machine Translation, UCT Journal of Social Sciences and Humanities Research 5(3) (2017) 89–100.