IEEE Turkey Section

UBMK'2021

Bildiriler Kitabı
Proceedings

6. Uluslararası Bilgisayar Bilimleri ve Mühendisliği Konferansı

6th International Conference on Computer Science and Engineering

15-16-17 Eylül (September) 2021 Ankara- Turkey

Finite State Machine Model for Uzbek Language Morphological Analyzer

Khamroeva Shahlo Mirdjanovna
Tashkent State University of Uzbek language and literature
Tashkent Uzbekistan
hamroyeva81@mail.ru

Abstract. **In this article, we discuss the development of a model of the Uzbek language FST (finite state transducer) in the creation of a morphological analyzer of the Uzbek language. There are key factors for automatic morphological analysis, such as stem, base, prefix, suffix, spelling rules. To do this, you need to create a database of word-formers in Uzbek (pre-/post-stem), lexical and syntactic suffixes, particles in the form of suffixes.**

Keywords: stem, base, prefix, suffix, spelling rules, Uzbek language FST

I. INTRODUCTION

For Turkic languages, which are included in agglutinative languages, the automatic approach of creating a morphological analyzer is more appropriate. The automated approach-based morphoanalyzer has a processing system FST (finite state transducer) and WFST (weighted finite state transducer), which can perform input and output (analysis/synthesis) analysis. The essence of such analyzers is that they follow the rule of "grammatical sequence": the process is based on the rule of morphological unity of the word. They differ in what units are used:

1) show the sequence of morphemes and the required allomorphic sequence;
2) rules of sequence of allomorphs.

In constructing a morphological analyzer of natural language, many scientists say [1] that it is an advantage to use the Finish State Machine method. In this article, we discuss the development of a model of the Uzbek language FST in the creation of a morphological analyzer of the Uzbek language.

There are key factors for automatic morphological analysis, such as stem, base, prefix, suffix, spelling rules. To do this, you need to create a database of word-formers in Uzbek (pre-/post-stem), lexical and syntactic suffixes, particles in the form of suffixes.

If all of this forms the database of the morphological analyzer, it is necessary to run the morphological analyzer and develop an analysis model to develop the program. Hence, for the morphological analysis of the word “daraxtlar” at the initial stage, the following information is required to be in the database:

1) Base (information indicating to which category the basis belongs);
2) Suffix (information representing the type).

Based on this information, the word form of “daraxtlar” are analyzed as follows: daraxt - ; -s is the plural form. When Adalı wrote about the creation of a morphological analyzer of Turkish and English, he distinguished the following database for the analyzer, which can identify the series of stems, additional and grammatical interpretation of the word [1]:

1) dictionary;
2) a series of suffixes;
3) spelling rules.

To do this, it is determined what suffixes a group of words can take. In Table 1 we present the suffixes that can be associated with a word and their position.

TABLE-I. WORD-FORMATIVE, POSSESSIVE SUFFIXES AND PREPOSITIONS ADDED TO THE UZBEK NOUN PHRASE

Word formative		**Prepositions**		**Possessive suffixes**	
singular	plural	singular	plural	singular	plural
gul	gullar	gul	gullar	gulim	gullari
gulzor	gulzorlar	gulning	gullarning	guling	gullaring
guldon	guldonlar	gulni	gullarni	guli	gullari
gulchi	gulchilar	gulga	gullarga	gulzorim	gulzorimiz
gulchilik	–	gulda	gullarda	gulzoring	gulzoringiz
gulfurush	gulfurushlar	guldan	gullardan	gulzori	gulzorlari

II. LITERATURE REVIEW

In Turkic computational linguistics, computer processing of a natural language, including the construction of morphoanalyzers, has been studied in detail. Among such works, we can note the works devoted to the creation of a morphological analyzer by E. Adali [1], K. Oflazer [10], A. Dybo, A. Sheimovich [2], Zh. Suleimanov, A.Gatiatullin, A.Bashirov [12], P. Zheltov [11], N. Israilova, P. Bakasova [7], N. Leontiev [9], V. Kukanova, A. Kadzhieva [8]. Although the development of a morphological analyzer for the Uzbek language has not been the subject of special studies, some studies have commented on some aspects of the problem [2].

While writing about the development of a morphological analyzer for languages belonging to different families, Adalı focuses on the problem of grammatical homonymy in the morphological analysis of agglutinative languages. It suggests considering different approaches to the definition of grammatical homonymy in agglutinative languages. In his opinion, it is necessary to develop different methods of morphological analysis of the word.

When creating a morphoanalyzer for any language, special attention is paid to the peculiarities of the language. Experts describe the Tatar language as a semantically complex, coherent, almost automatic linguistic phenomenon. Based on the study of the morphoanalyzer of the Tatar language, we came to the conclusion that the language is of practical importance in the process of searching and processing information of the Search information system. The main task of the MA of the Chuvash language is to determine the morphemic structure, morphological features of Chuvash words for syntactic and semantic analysis. The Chuvash MA database contains a list of types describing morphological features, a dictionary, reference, base and suffix database, text / word storage, working model structure data containing a sample of word analysis, general laws of word structure related to various categories of words. The morphoanalyzer of the Kyrgyz and Yakut languages has a similar component with some peculiarities. The experience in creating morphoanalysers of the Turkic languages is useful in developing the morphoanalyzer of the Uzbek language. Although the morphoanalyzer of the Uzbek language has not

been studied in monographic terms, there are works on some issues of morphological automatic analysis. For example, MA construction tools include vocabulary, morphotactics, spelling rules, information structure units that form the basic semantics. The study of the structure presented in the above works is useful in constructing a morphological analyzer of the Uzbek language.

III. FSM MODEL OF THE UZBEK LANGUAGE

"At the end of a morphological analysis of a word, one result may not be obtained, in other words, several results may emerge," the scholar writes and gives a few examples of grammatical homonymy [1]. The Uzbek language, which belongs to the family of agglutinative languages, also has this problem. Below we analyze the grammatical homonyms, the meaning of which in Uzbek is understood only in the context:

1.a. oting = ot+ing (name of the person: noun, general case, I person, singular);

1.b. oting = ot+ing (mammal: noun, general cas, I person, singular);

1.v. oting = ot+ing (to move an object from one starting point to another: verb, II person, plural, future tense, order speech, active voice).

2.a. tering = teri+ng (The outer covering of the human and animal body, the skin that protects the body parts from external influences and performs various physiological functions; general case, II person, singular)

2.b. tering = ter+ing (a clear fluid secreted by the sweat glands; general case, II person, singular);

2.v. tering = ter+ing (1. To pick up and collect spilled or spilled items one by one by hand; 2. Collecting grain, fruits, flowers and plant crops one by one; verb, II person, plural, future tense, order speech, exact ratio).

Which of these possible outcomes is correct can only be determined by context. This process is called morphological homonymy differentiation. In the future, this topic will be considered specifically, solutions will be proposed.

It is designed to show boundary and transition states in processes involving the FST. FST and its special forms are widely used in the construction of various linguistic analyzers. To understand the FST, we need to know its basic terms.

1. Initial state: Indicates the initial state of the limited state machine. It is indicated by an arrow coming from an unknown place.
2. Acceptance status: A status indicating that the FST has successfully completed its mission.
3. Recipient and Recognizers: Indicates whether the application has been accepted or not.
4. Transition motion: the transition from one state to another.
5. Converter: forms the output state using the input and motion used. There are two types of converters: (a) FST uses only input motions; the output depends on the situation; (b) (FST) uses only input actions: output depends on input and status.

A. Spelling Rules

Spelling rules also play an important role in creating a morphological analyzer. Although Uzbek is an agglutinative language among Turkic languages, there are some cases of inflection. Flexion cases occur as sound changes. For the analysis of word forms on the basis of the morphological analyzer should reflect the typical cases of sound changes, all sound changes in the Uzbek language. There are three main types of sound changes in the Uzbek language:

Sound Changes

The vowel at the end of a word changes with the addition of a suffix:

1) When the suffix -v, -q, -qi is added to verbs ending in a vowel, the vowel a is pronounced as o and written as: *sayla – saylov, sina – sinov, aya – ayovsiz; so'ra – so'roq, bo'ya – bo'yoq; o'yna – o'ynoqi, saura – sayroqi;*
2) When most verbs ending in the vowel i are followed by the suffix -v, -q, the vowel is pronounced and written as *u: o'qi – o'quvchi, qazi – qazuvchi, sovi – sovuq.* However, in some verbs ending in the vowel i, when the suffix -q is added, the vowel i is pronounced and written as *i: og'ri – og'riq, qavi – qaviq.*

When the possessive suffix is added to multi-syllable words ending in k, q, as well as to certain syllables such as bek, yo'q, the consonant k becomes the consonant g, the consonant q becomes the consonant g', and is written as follows: *tilak – tilaging, yurak – yuragim, kubok – kubogi, bek – begi; tayoq – tayog'i, qoshiq – qoshig'i, yaxshiroq – yaxshirog'i, yo'q – yo'g'i.* But in multi-syllable mastery words, when the possessive suffix is added to single-syllable plural words, the sound k, q is actually pronounced and written: *ishtirok – ishtiroki, ocherk – ocherki, erk – erki, huquq – huquqim, ravnaq – ravnaqi, yuq – yuqi.*

Sound Drop

With the addition of the following suffixes, the sound in the word structure drops:

1. When the possessive suffix is added to some words, such as *o'rin, qorin, burun, o'g'il, bo'yin, ko'ngil, and the suffix -il*, which forms the relative form to verbs such as *qayir, ayir, is added to ikki, olti, yetti words -ov,* The vowel in the second syllable is not pronounced or written when the suffix -ala is added: *o'rin – o'rnim, qorin – qorni, burun – burning, o'g'il – o'g'ling, ko'ngil – ko'ngli, yarim – yarmi; qayir – qayril, ulug' – ulg'ay, sariq – sarg'ay, ikki – ikkov, ikki – ikkala, yetti – yettov;*

Sound Increase

With the addition of the following suffixes, the sound content of the word increases:

2. It is pronounced with the sound n when the suffixes *-da, -dan, -day, -dagi, -ga, -gacha, -cha* are added to those pronouns *u, bu, shu, o'sha*, and it is written as follows: *unda, bunday, shunda, o'shancha;* the possessive suffixes to these pronouns are added as follows; *buningiz, o'shanisi*;
3. Possessive suffixes are added to words ending in o, o, u, e vowels as follows:

a) A lot of possessive suffixes *-m, -ng, -si; -miz, -ngiz, -si (or -lari)* are added without sound*: bobom, bobong, bobosi, bobomiz, bobongiz, bobosi (yoki bobolari); orzum, orzung, orzusi; orzumiz, orzungiz, orzusi;*

b) When the first and second person possessive suffixes are added to the words *parvo, obro, mavqe, mavzu, avzo,* a vowel is added and written as follows: *parvoyim, parvoying; parvoyimiz, parvoyingiz; obro'yim, obro'ying; obro'yimiz, obro'yingiz;* The possessive suffix of the third person is added to the words *parvo, avzo, obro', mavqe* in the form *-yi*, and to the words *xudo, mavzu* in the form of *-si*: as *avzoyi,*

mavzusi (as dohiy, -si is added to the word ending a "-y" consonant in the third person: like a dohiysi);

4. When the suffixes *-ni, -ning, -niki* are added to the pronouns *men, sen*, the sound n in the suffix is not pronounced or written: as *meni, mening, meniki; seni, sening, seniki.*

B. *Database of morphemes*

Another component required to run a morphological analyzer is a database of morphemes. Additions in the Uzbek language are divided into three groups in terms of function:

1) Word formers.
2) Syntactic formers.
3) Lexic formers.

To distinguish word-forming suffixes in the process of morphological analysis, a complete list of them should be provided. The list of word-formative suffixes that are actively used in the Uzbek language are given in Table-II.

These suffixes come in handy in distinguishing word-forming suffixes from other types of suffixes in the process of morphological analysis. Various lexical and syntactic suffixes are also actively used in the Uzbek language.

TABLE II. NOUN FORM SUFFIXES

Noun - noun form suffixes				
	suffix	**teg**	**explanation**	**example**
1	-bin:	Ŷ 1	noun - noun	Folbin
2	-bon:	Ŷ 1	noun - noun	darvozabon, soyabon, tarozibon, xazinabon
3	-boz:	Ŷ 3	noun - noun	masxaraboz, qimorboz, dorboz
4	-voy:	Ŷ 4	noun - noun	novvoy (nonvoy)
5	-gar//-kar:	Ŷ 5	noun - noun	zargar, savdogar, da'vogar, miskar
6	-garchilik:	Ŷ 6	noun - noun	yog'ingarchilik, odamgarchilik
7	-gin	Ŷ 7	noun - noun	jahongir, fazogir
8	-goh:	Ŷ 8	noun - noun	oromgoh, saylgoh, sayrgoh, qarorgoh, ziyoratgoh, bazmgoh
9	-go'y:	Ŷ 9	noun - noun	kalimago'y, maslahatgo'y
10	-diq//-dik	Ŷ 10	noun - noun	o'rindiq
11	-don:	Ŷ 11	noun - noun	guldon, kuldon, qalamdon
12	-don	Ŷ 12	noun - noun	muhrdor, chorvador
13	-dosh:	Ŷ 13	noun - noun	sinfdosh, kursdosh, maslakdosh
14	-do'z:	Ŷ 14	noun - noun	etikdo'z, mahsido'z, kashtado'z
15	-zor:	Ŷ 15	noun - noun	olmazor, gulzor, olchazor
16	-iston:	Ŷ 16	noun - noun	guliston, go'riston, O'zbekiston
17	-kash:	Ŷ 17	noun - noun	aravakash, qalamkash, suratkash
18	-kor:	Ŷ 18	noun - noun	ganchkor, paxtakor, sholikor, san'atkor
19	-kov:	Ŷ 19	noun - noun	go'rkov
20	-lik/liq:	Ŷ 20	noun - noun	bolalik, vaqtichog'lik, do'stlik, boshliq
21	-loq:	Ŷ 22	noun - noun	O'tloq, qumloq, toshloq
22	-noma:	Ŷ 23	noun - noun	taklifnoma, tabriknoma, pandnoma
23	-navis:	Ŷ 24	noun - noun	tarixnavis, voqeanavis, romannavis
24	-paz	Ŷ 25	noun - noun	oshpaz, kabobpaz, somapaz
25	-soz	Ŷ 26	noun - noun	soatsoz, kemasoz
26	-fuwsh:	Ŷ 27	noun - noun	baliqfurush, nosfurush
27	-xon:	Ŷ 28	noun - noun	kitobxon, she'rxon
28	-xona:	Ŷ 29	noun - noun	darsxona, mehmonxona, ishxona,
29	-xo'r	Ŷ 30	noun - noun	merosxo'r
30	-cha	Ŷ 31	noun - noun	qalamcha
31	-chak//-choq:	Ŷ 32	noun - noun	o'yinchoq
32	-chi:	Ŷ 33	noun - noun	ishchi, temirchi, terimchi, gulchi, bosqinchi
33	-chilik	Ŷ 34	noun - noun	hunarmandchilik, o'zbekchilik, dehqonchilik
34	Ham-	Ŷ 35	noun - noun	hamqishloq, hamshahar, hamyurt
Noun formers from the base of another word category				
35	-a (	Ŷ 36	imitation	qahqah, sharshar, g'arg'ar, jizz,
36		Ŷ 37	adjectıve - noun	bo'z, quyuq
37	-a	Ŷ 38	adjectıve - noun	xarob, vayron
38	-ak:	Ŷ 39	imitation	bizbiz, pirpir, guldir, var, qar, xur
39	-archilik:	Ŷ 40	adjectıve - noun	Och
40	-at:	Ŷ 41	verb - noun	ko'chat, o'lat
41		Ŷ 42	adjectıve - noun	ko'kat
42	-vchi//-uvchi:	Ŷ 43	verb - noun	o'quvchi, yozuvchi, uchuvchi, aniqlovchi, to'kjiruvchi
43	-garchilik	Ŷ 44	adjectıve - noun	namgarchilik, xafagarchiiik, xunobgarchilik, sharmandagarchilik
44	-gi//-ki// -qi//-g'i//-g'u:	Ŷ 45	verb - noun	sezgi, sevgi, supurgi, kukju, turtki sanchqi, tomizg'i, tuyg'u.
45	-gach//-kich// -qich//-g'ich:	Ŷ 46	verb - noun	kulgich, o'tkazgich, ko'rsatkich, yoritqich, tutqich, ochqich, to'g'nag'ich, chizg'ich, o'chirg'ich
46	-gin//-qin// -kin//-gun// -qun:	Ŷ 47	verb - noun	tizgin, surgun, tolqin, to'sqin, quvg'in, yong'in, uchqun
47	-dak//-doq:	Ŷ 48	verb - noun	yugurdak, kekirdak, qovurdoq, qo'ndoq
48	-diq	Ŷ 49	verb - noun	qoldiq, topildiq, hordiq
49	-ik:	Ŷ 50 (1)	verb - noun	ko'rik, teshik (noun - noun va adjectıve - noun), kekirik
50	-ik:	Ŷ 50(2)	verb - noun	bilik, bitik
51	-ik:	Ŷ 50(3)	adverb - noun	ko'pik
52	-ildoq:	Ŷ 51	imitation	hiqildoq, chirildoq
53	-imlik:	Ŷ 52	verb - noun	o'simlik, ichimlik
54	-in//-un:	Ŷ 53	verb - noun	yig'in, yog'in, ekin, tiqin, tugun, tutun
55	-indi// undi//ndi:	Ŷ 54	verb - noun	chiqindi, yuvundi, chirindi, cho'kindi, kuyundi, yig'indi
56	-it:	Ŷ 55	verb - noun	chiqit
57	-ich:	Ŷ 56	verb - noun	cho'mich, cho'kich, o'pich, bog'ich
58	-ish:	Ŷ 57	verb - noun	Qarg'ish
59	-iq//-uq:	Ŷ 58	verb - noun	chaqiriq, kesatiq, yutqiziq, chopiq, yutuq
60	-k:	Ŷ 59	verb - noun	ko'rik, elak, tilak, kurak, bezak, to'shak
61	-kilik//-gilik:	Ŷ 60	verb - noun	ichkilik, ko'rgilik
62	-lik	Ŷ 61(1)	numeratıv-noun	birlik, to'rtlik
63		Ŷ 61(2)	Pronoun - noun	o'zlik
64		Ŷ 61(3)	adverb - noun	tezlik, sekinlik, birgalik
65		Ŷ 61(4)	Modal - noun	borliq, yo'qlik
66	-m//-im//-um:	Ŷ 62(1)	verb - noun	to'plam, ho'plam, chidam, tishlam, kechirim, qo'nim, terim, chiqim, bitim, bosim, unum, tuzum
67		Ŷ 62(2)	imitation	qultum
68	-ma:	Ŷ 63	verb - noun	surma, o'sma, tortma
69	-mak//-moq:	Ŷ 64	verb - noun	yemak, ilmoq, chaqmoq, topishmoq
70	-machoq:	Ŷ 65	verb - noun	bekinmachoq, quvlashmachoq
71	-mish:	Ŷ 66	verb - noun	noun - nounmish, kechmish, qilmish.
72	-movchilik:	Ŷ 67	verb - noun	anglashilmovchiiik, kelishmovchilik, yetishmovchilik
73	-on:	Ŷ 68	verb - noun	qiron, to'zon
74	-os:	Ŷ 69	imitation	uvvos, chuvvos
75	-ot	Ŷ 70	Adjective-noun	ma'lumot, mushkulot xarobot
76	-at:	Ŷ 71	adjectıve - noun	she'riyat, madaniyat, majburiyat
77	-ch//-j//-inch:	Ŷ 72	verb - noun	sevinch, quvonch, yupanch, ilinj, qo'rqinch
78	-cha:	Ŷ 73(1)	verb - noun	tushuncha
79	-cha:	Ŷ 73 (2)	adjectıve - noun	qizilcha, olacha
80	-chak	Ŷ 74	verb - noun	belanchak, ovunchoq, taqinchoq
81	-chi	Ŷ 75(1)	adjectıve - noun	qiziqchi
82	-chi	Ŷ 75(2)	verb - noun	suyunchi, tilanchi, tomchi
83	-chilik:	Ŷ 76(1)	undov	haybarakallachi
84	-chilik:	Ŷ 76(2)	adjectıve - noun	pishiqchiiik, arzonchilik
85	-chiq:	Ŷ 77	adverb - noun	ko'pchilik, ozchilik
86	-shunos:	Ŷ 78	verb - noun	suyanchiq, yopinchiq

IV. ABOUT UZBEK FST MODELS

C. *Morphological analysis of word group noun*

Word-formation, lexical and syntactic suffixes are added to the nouns in the Uzbek language. In the Uzbek language, noun-forming suffixes from different stems. In the Uzbek language, the forms of words in the noun category are usually arranged in the following sequence: Base + word-former + number + possession + case. These suffixes are necessary in the analysis of the Uzbek language horse phrase based on FST models. The rules of spelling formed because of the addition of a horse are given in spelling rules. The FST model of the analysis based on these appendices is given in Figure 1.

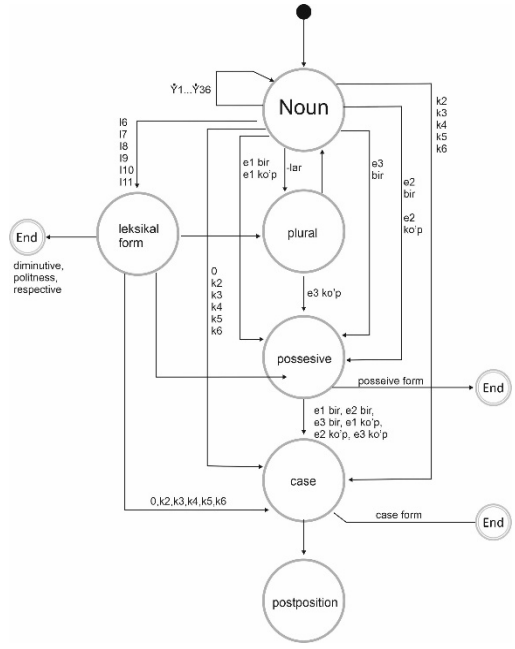

Fig 1: FST model of the noun

The first introductory part of the analysis is marked with a black dot. To make the model clear, the tags shown in Table-II and III of the appendices are written between actions. When

suffixes are involved in the transition from one grammatical form to another, they are indicated by an arrow, cases of non-indicative change of grammatical meaning are indicated by an empty arrow. We explain the diagram below.

Noun is the base: The condition indicated by a black dot indicates a word coming from outside for analysis. This provides access to the FSM. It is considered an artificial noun with Ŷ1 ... Ŷ35 suffixes in the base. Even if a noun-former comes after the base, the noun is made. For example: *fol+bin, darvoza+bon, soya+bon, zar+gar, savdo+gar, nam+garchilik, xafa+garchilik, sez+gi, sev+gi.* The drawing shows that a new noun was made from the noun's core. The noun can be in a state of transition from the base to several forms. These can be noun cases in the plural, lexical form, possessive, and accusative forms.

Plural: The form base+lar makes the plural forms of the noun: *bola+lar, olma+lar.* The word can end in this form or take the forms of possession and agreement, so there are three exits from this form:

1) plural noun: qizlar, gullar;
2) a plural noun that takes the form of a case: *qizlarlarning, qizlarni, qizlarga, qizlarda, qizlardan;*
3) noun in possession and plural: *qizlarim, qizlaring, qizlari, qizlarimiz, qizlaringiz, qizlari.*

Possession: status e1bir, e2bir, e3bir, e1kop, e2kop, e3kop take suffixes have many symbols and have two outputs:

1) A noun in the form of possession: *kitobim, kitobing, kitobi, kitobimiz, kitobingiz, kitoblari;*
2) A noun in the form of possession and case: *kitobim, kitobimning, kitobimni, kitobimga, kitobimda, kitobimdan.*

The state of the lexical form of the noun: the lexical form of the noun is formed by suffixes marked with tags I6, I7, I8, I9, I10, I11. For example: *odamniki, tog‘agi, hovligacha, bo‘taloq, onaxon, ukajon.* There will be 3 exit points from this case:

1) The lexical form of the noun: *uydagi, opajon*;
2) The lexical form of the possessive noun: uyimizdagi, onaxonimiz;
3) The lexical form of the plural noun: *uydagini, onajonimni.*

The state of the noun in the **form of a case** is formed in the following conditions.

The first condition: A direct case suffix is added to the base: base + 0 / k2 / k3 / k4 / k5 / k6. For example: *xona, xonaning, xonani, xonaga, xonada, xonadan.*

The second condition: To the base that takes the lexical form, an case suffix is added: base + I6 / I7 / I8 / I9 / I10 / I11 + 0 / k2 / k3 / k4 / k5 / k6. For example: *bolajon, bolajonning, bolajonni, bolajonga, bolajonda, bolajondan.*

The third condition: A case suffix is added to the base that takes the form of possession: base + e1bir / e2bir / e3bir / e1kop / e2kop / e3kop + 0 / k2 / k3 / k4 / k5 / k6. For example: *kitobim, kitobimning, kitobimni, kitobimga, kitobimda, kitobimdan.*

There can be two output states from the contract form. A noun in the form of a case and a noun that receives an auxiliary after the contract.

D. Model of morphological analysis of verb phrases

In the Uzbek language, suffixes are added to the verb base in the order to word-former + lexical form-former + syntactic form-maker. In Uzbek, a compound verb is formed by adding an adjective to a non-verb base.

We have given these additions that are needed in the analysis based on the Uzbek verb phrase FST models in the tables above. The rules of spelling formed because of the addition of a verb are given in Figure 2 and.3. The FSTM model of the verb, which can be analyzed based on these suffixes database and spelling rules, is given in Figure 2.

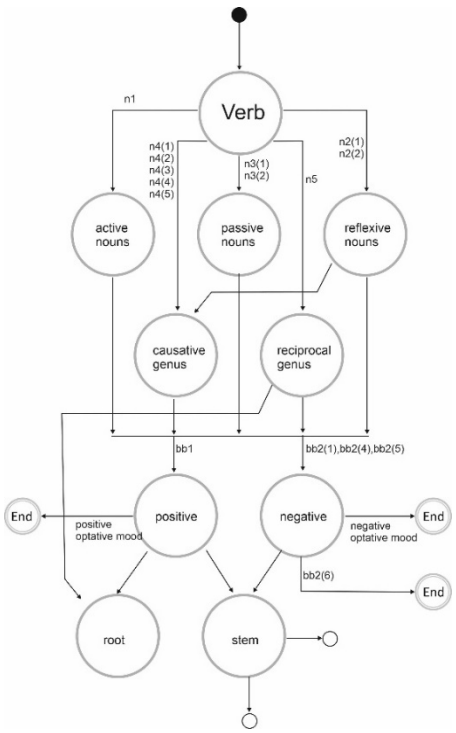

Fig 2: FST model of the verbs

We interpret the diagram as follows. The black dot indicates the entry status of the verb stem into the analysis. There are 5 transition modes for switching from verb base to relative suffixes.

The first case: When n1 is added to the verb stem / base, an active voice is formed. There is only one way out of this situation: 1) a state that takes the form of affirmative / negative: *kel/kelma, bor/borma, o‘tir/o‘tirma, yoz/yozma*.

The second case: When n2 (1), n2 (2) forms are added to the base of the verb, the personal pronoun is formed: *yuvin(di), taran(di), bezan(di), yozil(di), ushlan(di).* There are two transition states from this state: 1) superlative form; 2) Switch to affirmative / negative form. When the form bb1 is added, the participle is formed, bb2 (1), bb2 (4), bb2 (5) is formed without the participle: *yuvin(di)-yuvinma(di), taran(di) - taranma(di), bezan(di) - bezanma(di).*

Third case: When n3 (1), n3 (2) are added to the base of the verb, a passive ratio is formed: *yuvil(di), taral(di), bezal(di), yozil(di), ushlan(di).* There are two transition states from this state. Affirmative or negative form: affirmative verb form is formed when bb1 is added, bb2 (1), bb2 (4), bb2 (5) is formed without negative.

Fourth case: When n5 is added to the stem \ base, a compound voice is formed: *o‘tirish(di), yozish(di), boshlash(di).* There are two transition states from this state: the transition to the affirmative or negative form. The participle form bb1 (1), bb2 (4), and bb2 (5) are formed when the form bb1 is added.

Fifth case: When n4 (1), n4 (2), n4 (3), n4 (4), n4 (5) are added to the base of the verb, an plural form is formed: *chiqar(di), tomiz(di), yuvdir(di), o‘qit(di), keltir(di), o‘tqaz(di), o‘tkaz(di), oshir(di).* There are two transition states from this state. A form of affirmative or negative. When the form bb1 is added, the verb to be affirmative, when the form bb2 (1), bb2 (4), bb2 (5) is added negative form is made: *chiqar(di) - chiqarma(di), tomiz(di) - tomizma(di), yuvdir(di) - yuvdirma(di), o‘qit(di) - o‘qit(di), keltir(di) - keltir(di), o‘tqaz(di), o‘tkaz(di) o‘tkaz(di), oshir(di) - oshir(di).*

Verb word form (verb / stem). *kel, o‘tir, yoz, yuvin, tugat, asra, o‘tin.* There are three exit states from the verb form without negative. imperative order-request affirmative verb: *borma, kelma, o‘tirma, yozma, yuvinma, tugatma, asrama, o‘tinma*. The form of negative action name: *kelmaslik, o‘tirmaslik, yozmaslik, yuvinmaslik, tugatmaslik, asramaslik, o‘tinmaslik. V*erb stem (verb word form): *borma, kelma, o‘tirma, yozma, yuvinma, tugatma, asrama, o‘tinma.*

It is possible to distinguish the forms of “mood” from a single exit state of the verb form shown in Fig. 3: indicative mood, jussive mood, conditional mood.

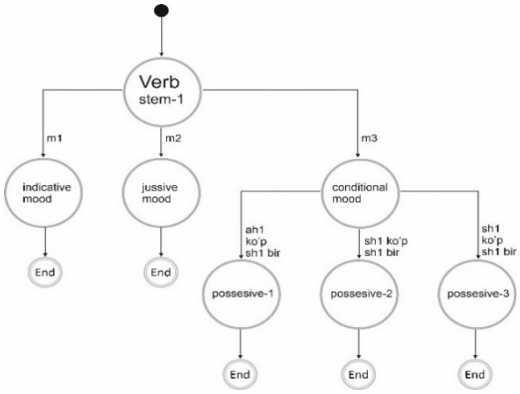

Fig 3: FST model of single exit state of the verb form

1) if the verb stem + m1 = indicative mood. For example: bordim, bording, bordi.
2) if the verb stem + m2 = jussive mood. For example: boray, bor(gin), borsin, boraylik, boringiz, borsinlar; there is a single output (last case) from the jussive mood.
3) if the verb stem + m3 = conditional mood. For example: borsam, borsang, borsa, borsak, borsangiz, borsalar. There are three exits (last case) from the conditional mood. I person singular / plural conditional mood: borsam, borsak; II person singular / plural conditional mood: borsang, borsangiz; III person singular / plural conditional mood: borsa, borsalar.

From the second form of the verb to 3 states: 1) the function form of the verb; 2) the tense form of the verb; 3) a form of auxiliary verb that expresses the strength / weakness of an action.

1) If verb-stem + vhn1\ vhn2 \ vhn3, the masdar is generated: o‘qimoq, o‘qish, o‘quv; yozmoq, yozish, yozuv; there is only one output mode.
2) if the verb-stem + vsif1 \ vsif2 \ vsif3, a participle is formed: borgan, borayotgan, boradigan, o‘qigan, o‘qiyotgan, o‘qiydigan; there is only one output mode.
3) if the verb-stem + vrav1 \ vrav2 \ vrav3 \ vrav4 \ vrav5 \ vrav6, converb is formed: borib, o‘qib, borgancha, to‘kkancha, siqqancha, borgach, kelgal, to‘kkach, borguncha, tikkuncha, yiqquncha, borgani, to‘kkani, kelgali; there is only one output mode.

The second transition state from the verb form indicates the strong / weak action and the auxiliary verb form (Fig 4).

1) If the verb-stem + hks1 \ hks2 \ hks3, a verb form representing the weakness of the action is formed. For example: yig‘lamsira(di), oqarinqira(di), to‘lish(di);
2) If the verb-stem + hks11 \ hks12 \ hks13 hks11 \ hks12 \ hks13, hks14 \ hks15 \ hks16 \ hks17, a verb form representing the strength of the action is formed. For example: tepkila, turtkila, chayqa;
3) If the verb-stem + h.kmf1… h.kmf34, an auxiliary verb conjugation is formed. For example: yozaver, aytaqol, o‘qiyver.

There is also a transition from verb form to tense. There will be 3 output states from the time form.

1) Past tense form: The verb form with the suffixes Z 1(1)… Z1(16) is considered past tense. From this form there is one exit state, one person-number transition state. If the form Z 1(1) is added to the verb stem, the past tense verb is formed: keldi, bordi, o‘tirdi.

If the forms Z 1(2)… Z1 (16) are added, it is converted to the person-number form and has 3 outputs.

First person: *bordim, borganman, boribman; bordik, borganmiz, boribmiz;*

The second person: *bording, borgansan, boribsan; bordingiz, borgansiz, boribsiz;*

Third person: *bordi, borgan, boribdi; bordilar, borganlar, boribdilar.*

2) Present tense form: Z2(1)… Z2(18). From this form there is one exit state and one person-number transition state. If the form Z2(1) is added to the verb stem, the present tense verb is formed: kelyapti, bormoqda, o‘tiryapti. If the form Z2(4) ... Z2(18) is added to the verb stem, the person-number mode is changed:

First person: boryapman, bormoqdaman; boryapmiz, bormoqdamiz;

The second person: boryapsan, bormoqdasan, boryapsiz, bormoqdasiz;

Third person: boryapti, bormoqda; boryaptilar, bormoqdalar.

3) Future tense form: Z3(1)… Z3(18). From this form there is one exit state and one person-number transition state. If Z3 (4), Z3 (7), Z3 (8), Z3 (9), Z3 (10), Z3 (14) are added to the verb stem, the future tense verb is formed and the output is as follows: -r / ar: -r emish / ar emish: // - rmish / armish: -r ekan / ar ekan: // - rkan / arkan: -moqishi: moqchi emish: -gay / kay / qay. For example: o‘qir, yozar, borarmish, borar ekan, borarkan, bormoqchi, borgay.

If the forms Z3 (1), Z3 (4), Z3 (5), Z3 (6), Z3 (7), Z3 (8), Z3 (9), Z3 (10), Z3 (14), Z3 (15) are added to verb stem, the transition from the verb of the future tense to the person-number is made and it has 3 outputs: -y/ a; -r/ ar; -r emish/ ar emish; -r ekan/ ar ekan; -moqchi; -moqchi emish; -moqchi ekan; -yajak/ ajak; -gay/ kay/ qay; -gu/ ku/ qu.

The first person: boraman, borarman, borar emishman, borajakman, borgayman; boramiz, borarmiz, borar emishmiz, borajakmiz, borgaymiz;

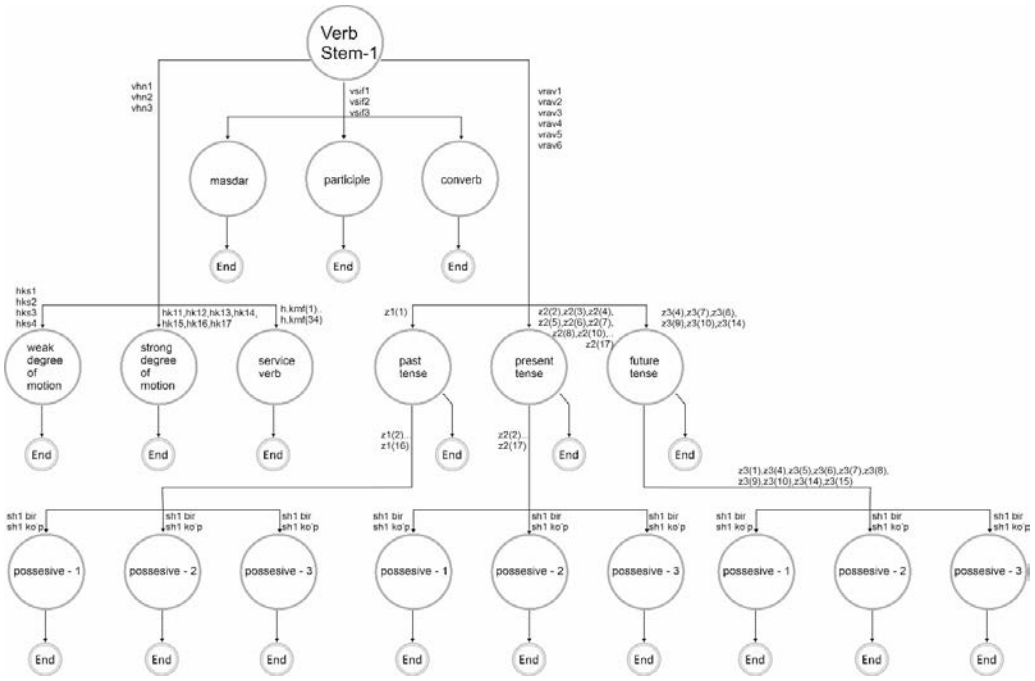

Fig 4: FST model of single exit state of the verb form

The second person: *borasan, borarsan, borar emishsan, borajaksan, borgaysan; borasiz, borarsiz, borar emishsiz, borajaksiz, borgaysiz;*

Third person: *boradi, borar, borar emish, borajak, borgay; boradilar, borarlar, borar emishlar, borajaklar, borgaylar.*

V. Conclusion And Recommendations

In conclusion, we can state, that the FST model performs morphological analysis from left to right. It should be noted that these developed FST models will serve as a linguistic support for the development of a morphological analyzer of the Uzbek language in the future. The database and tags of Uzbek grammatical forms also serve as linguistic support in the implementation of these models. So, to create a morphological analyzer of the Uzbek language, you need to perform the following algorithm:

1) compiling a dictionary of the basics of the Uzbek language;

2) formation of the database of grammatical forms of the Uzbek language and designation of their tags;

3) entering a set of orthographic rules of the Uzbek language into the database;

4) development of FST models of morphological analysis of words.

The linguistic support of the morphoanalyzer is stored in the form of a database: the database consists of a dictionary of basics and a dictionary of rules. A database is a collection of extensive data in which data is sorted and stored in tabular form in a database system. The dictionary of bases consists of the lexicon of the language, the grammatical dictionary compiling a list of grammatical forms in the language; base of affixes; classification rules; allomorph compatibility rules are located in the database. Linguistic support of morphoanalyzer reflects phonetic changes in the Uzbek language at the core and additional boundaries; include information about suppletive, invariant lexemes, along with conventional syntactic form-makers; embodies the base of homonymous lexemes.

References

[1] E. Adalı "Türkçe Doğal Dil İşleme", Akçağ Yayınev, 2020.

[2] T.B. Boltaev, S.I. Ibragimov "On the project of a software system for morphological analysis of the Uzbek language" // http://buxdu.uz/index.php/uz/

[3] A.V. Dybo , A.V. Sheimovich "Automatic morphological analysis for the corpora of the Khakass and ancient Turkic languages", Scientific Review of the Sayano-Altai peer-reviewed scientific journal No. 2 (08), 2014

[4] R.R. Gataullin, A.R. Gatiatullin, O.A. Nevrozova, D.R. Muxamedshin, D.Sh Suleymanov, B.E. Xakimov, A. F. Xusainov "Formalnie modeli i programnie instrumenti kompyuternoy obrabotki tatarskogo yazika"– Akademiya nauk RT, Institut priklalnoy semiotiki AN RT. – Kazan: Akademii nauk, 2019. – 260 s. – S.39-40.

[5] A.R. Gatiatullin, A. M. Bashirov, "Morphological analyzer of Turkic word forms based on the structural-functional model of the Turkic morpheme", Fifth Int. Conf. on Computer Processing of Turkic Languages "TurkLang 2017. Proceedings of the conf. vol-2 Kazan.

[6] A.F. Gelbukh, "Effectively realizable morphologic model of inflective language" Nauchno-texnicheskayaInformasiya series 2, #1, 1992. – R. 24-31.;

[7] N.A. Israilova, P.S. Bakasova,"Morphological analyzer of the Kyrgyz language", Fifth Int. Conf. on Computer Processing of Turkic Languages "TurkLang 2017". Proceedings of the conf. vol-2. Kazan

[8] V.V. Kukanova, A.Yu. Kadzhiev,"Algorithm of the morphological parser of the Kalmyk language", In the collection: Pismenoto heritage and information technologies. El'Manuscript – 2014 pp. 116-119.

[9] N.A. Leontiev,"Morphological analyzer of the Yakut language", Sixth Inter. Conf. on Computer Processing of Turkic Languages "TurkLang-2018". Tashkent, 2018. Pp:. 276-279.

[10] K. Oflazer, "Two-level Description of Turkish Morphology", Literary and Linguistic Computing, – Vol. 9, No 2, – 1994.;

[11] P.V. Zheltov, "Development of a morphological analyzer of the Chuvash language", Fifth Inter. Conf. on Computer Processing of Turkic Languages "TurkLang 2017" vol-2 Kazan

[12] D.Sh. Suleimanov, R.A. Gilmullin, R.R. Gataullin."Morphological analyzer of the Tatar language based on a two-level model of morphology" Fifth Inter. Conf. on Computer Processing of Turkic Languages "TurkLang 2017". Vol-2. – Kazan

[13] V.A. Tuzov "Morfologicheskiy analizator russkogo yazika" Vestnik SPbGU, ser. 1. 1996. Vip. 1 (N15). − S. 41-45.

[14] Ç. Çöltekin "A Set of Open Source Tools for Turkish Natural Language Processing In" : Proc. of the Ninth Inter. Conf. on Language Resources and Evaluation (LREC'14) Ed. by N. Calzolari et al. 1079-1086. 2014.