Rule Based Morphological Analyzer of Kazakh Language

Gulshat Kessikbayeva
Hacettepe University, Department of Computer Engineering,
Ankara,Turkey
shatik2030@gmail.com

Ilyas Cicekli
Hacettepe University, Department of Computer Engineering,
Ankara,Turkey
ilyas@cs.hacettepe.edu.tr

Abstract

Having a morphological analyzer is a very critical issue especially for NLP related tasks on agglutinative languages. This paper presents a detailed computational analysis of Kazakh language which is an agglutinative language. With a detailed analysis of Kazakh language morphology, the formalization of rules over all morphotactics of Kazakh language is worked out and a rule-based morphological analyzer is developed for Kazakh language. The morphological analyzer is constructed using two-level morphology approach with Xerox finite state tools and some implementation details of rule-based morphological analyzer have been presented in this paper.

1 Introduction

Kazakh language is a Turkic language which belongs to Kipchak branch of Ural-Altaic language family and it is spoken approximately by 8 million people. It is the official language of Kazakhstan and it has also speakers in Russia, China, Mongolia, Iran, Turkey, Afghanistan and Germany. It is closely related to other Turkic languages and there exists mutual intelligibility among them. Words in Kazakh language can be generated from root words recursively by adding proper suffixes. Thus, Kazakh language has agglutinative form and has vowel harmony property except for loan-words from other languages such as Russian, Persian and Arabic.

Having a morphological analyzer for an agglutinative language is a starting point for Natural Language Processing (NLP) related researches. An analysis of inflectional affixes of Kazakh language is studied within the work of a Kazakh segmentation system (Altenbek and Wang, 2010). A finite state approach for Kazakh nominals is presented (Kairakbay and Zaurbekov, 2013) and it only gives specific alternation rules without generalized forms of alternations. Here we present all generalized forms of all alternation rules. Moreover, many studies and researches have been done upon on morphological analysis of Turkic languages (Altintas and Cicekli, 2001; Oflazer, 1994; Coltekin, 2010; Tantug et al., 2006; Orhun et al, 2009). However there is no complete work which provides a detailed computational analysis of Kazakh language morphology and this paper tries to do that.

The organization of the rest of the paper is as follows. Next section gives a brief comparison of Kazakh language and Turkish morphologies. Section 3 presents Kazakh vowel and consonant harmony rules. Then, nouns with their inflections are presented in Section 4. Section 4 also presents morphotactic rules for nouns, pronouns, adjectives, adverbs and numerals. The detailed morphological structure of verbs is introduced in Section 5. Results of the performed tests are presented together with their analysis in Section 6. At last, conclusion and future work are described in Section 7.

2 Comparison of Closely Related Languages

There are many studies and researches prior made on closely related languages by comparing them for many purposes related with NLP such as Turkish–Crimean Tatar (Altintas and Cicekli, 2001), Turkish–Azerbaijani (Hamzaoğlu, 1993), Turkish–Turkmen (Tantuğ et al., 2007), Turkish-Uygur (Orhun et al, 2009) and Tatar-Kazakh (Salimzyanov et al, 2013). A deep comparison of Kazakh and Turkish languages from computational view is another study which is in out of scope for this work. However, in this study, a brief grammatical comparison of these languages is given in order to give a better analysis of Kazakh language.

Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM, pages 46–54,
Baltimore, Maryland USA, June 27 2014.

Kazakh and Turkish languages have many common parts due to being in same language family. Possible differences are mostly morpheme based rather than deep grammar differences. Distinct morphemes can be added in order to get same meaning. There exist some differences in their alphabets, their vowel and consonant harmony rules, their possessive forms of nouns, and inflections of verbs as given in Table 1. There are extra 9 letters in Kazakh alphabet, and Kazakh alphabet also has 4 additional letters for Russian loan words.

Both Kazakh language and Turkish employ vowel harmony rules when morphemes are added. Vowel harmony is defined according to last morpheme containing back or front vowel. In Kazakh language, if the last morpheme contains a back vowel then the vowel of next coming suffix is *a* or *ı*. If the last morpheme contains one of front vowels then the vowel of next coming suffix is *e* or *i*. In Turkish, suffixes with vowels *a*, *ı*, *u* follow morphemes with vowels *a*, *o*, *u*, *ı* and suffixes with vowels *e*, *i*, *ü*, *ö* follow morphemes with vowels *e*, *i*, *ü*, *ö* depending on being rounded and unrounded vowels. Consonant harmony rule related with voiceless letters is similar in both languages.

	Turkish	**Kazakh Language**
Alphabet	Latin 29 letters (8 Vowels, 21 Consonant)	Cyril 42 letters (10 Vowels, 25 Consonants, 3 Compound Letters, 4 Russian Loan Word Letters)
Vowel & Consonant Harmony	Synharmonism according to back, front, unrounded and rounded vowels	Synharmonism according to back and front vowels
Possessive Forms of Nouns	6 types of possessive agreements	8 types of possessive agreements
Case Forms of Nouns	7 Case Forms	7 Case Forms
Verbs	Similar Tenses	Similar Tenses

Table 1. Comparison of Kazakh and Turkish.

In Kazakh language there are 8 types of personal possessive agreement morphemes as given in Table 2. Kazakh language has two additional possessive agreements for second person.

There are some identical tenses and moods of verbs in both language such as definite past tense, present tense, imperative mood, optative mood and conditional mood. They have nearly same morphemes for tenses. On the other hand there are some tenses of verbs which are identical according to meaning and usage, but different morphemes are used. Moreover, in Kazakh language there are some tenses such as goal oriented future and present tenses which do not exist in Turkish language.

Possessive Pronoun	**Representation**		**Examples for Eke, "father"**	
None Possessive	Pnon		Eke	father
My	P1Sg	1	Eke-m	my father
Your	P2Sg	2	Eke-N	your father
Your (Polite)	P2PSg	2	Eke-Niz	your father
His/Her	P3Sg	3	Eke-si	his father
Our	P1Pl	1	Eke-miz	our father
Your Plural	P2Pl	2	Eke-leriN	your father
Your Plural (Polite)	P2PPl	2	Eke-leriNiz	your father
Their	P3Pl	3	Eke-leri	their father

Table 2. Possessive Agreement of Nouns.

3 Vowel and Consonant Harmony

Kazakh is officially written in the Cyrillic alphabet. In its history, it was represented by Arabic, Latin and Cyrillic letters. Nowadays switching back to Latin alphabets in 20 years is planned by the Kazakh government. In the beginning stage of study, Latin transcription of Cyril version is used for convenience.

Two main issues of language such as morphotactics and alternations can be dealt with Xerox tools. First of all, morphotactic rules are represented by encoding a finite-state network. Then, a finite-state transducer for alternations is constructed. Then, the formed network and the transducer are composed into a

single final network which cover all morphological aspects of the language such as morphemes, derivations, inflections, alternations and geminations (Beesley and Karttunen, 2003).

Vowel harmony of Kazakh language obeys a rule such that vowels in each syllable should match according to being front or back vowel. It is called synharmonism and it is basic linguistic structure of nearly all Turkic languages (Demirci, 2006). For example, a word *qa-la-lar-dIN,* "of cities" has a stem *qa-la*, "city" and two syllables of containing back vowels according to the vowel harmony rule. Here *–lar* is an affix of Plural form and *–dIN* is an affix of Genitive case. However, as stated before, there are a lot of loan words from Persian and generally they do not obey vowel harmony rules. For example, a word *mu-Ga-lim,* "teacher" has first two syllables have back vowels and the last one has a front vowel. So suffixes to be added are defined according to the last syllable. For example, a word *muGalim-der-diN*, "of teachers" has suffixes with front vowels. On the other hand, there are morphemes with static front vowels which are independently from the type of last syllable can be added to all words such as Instrumental suffix *–men.* In this case, all suffixes added after that should contain front vowels.

Name	XFST	Type 1	Type 2
Sonorous Consonant	SCons	**l r y w**	**m n N**
Voiced Consonant	VCons	**z Z**	**b v g d**
Voiceless Consonant	VLCons	**p f q k t s S C x c**	
Consonant	Cons	**b p t c x d r z Z s S C G f q k g N l m n h w y v**	
Vowel	Vowel	**a e E i I O o u U j**	
Front Vowel	FVowel	**e E i O U j**	
Back Vowel	BVowel	**a I o u**	

Table 3. Groups of Kazakh letters according to their sound. Upper case letters are used for non-Latin letters.

In order to construct a finite-state transducer for alternation rules, there are some capital letters such as *A*, *J*, *H*, *B*, *P*, *C*, *D*, *Q*, *K, T* are defined in intermediate level and they are invisible by user. These representations are used for substitution such as *A* is for *a* and *e* and *J* is for *I* and *i*. So if suffix *dA* should be added according to morphotactic rules, it means suffixes *da* or *de* should be considered. In Table 3, there are group of letters defined according to their sounds and these groups are used in alternation rules (Valyaeva, 2007).

Consonant harmony rules are varied according to the last letter of a word with in morphotactic rules. As in Table 3, different patterns are presented in order to visualize the relation between common valid rules and to generalize morphotactic rules. Thus, in each case according to morphotactic rules there are proper alternation rules for morphemes.

GROUP 1					
Ablative Case		Locative case		Dative Case	
dAn		dA		TA	
tAn		tA		TA	
nAn	3	ndA	3	nA	3
				A	1/2
GROUP 2					
Genitive Case		Accusative Case		Poss. Affix-2	
dJN		dJ		diki	
tJN		tJ		tiki	
nJN	3	nJ		niki	
		n	3		
GROUP 3					
Plural Form of Noun		Negative Form		A1Pl	
dAr	l	bA		bJz	
tAr		pA		pJz	
lAr	r y w	mA		mJz	
GROUP 4					
Instrumental Case				A1Sg	
ben				bJn	
pen				pJn	
men	3			mJn	

Table 4. Alternation rules according to groups of letters.

All alternation rules for suffixes depend on the last letter of a morpheme with in morphotactic rules and Table 4 gives some groupings that can be made in order to set some generalized rules overall. Patterns of last letters of morphemes in Table 4 are matched with groups of letters presented in Table 3. In Table 4, Locative case affix is *–dA*, if the last letter of a morpheme is one of Vowel, Sonorous

Consonant or Voiced Consonant of Type 1 in Table 3. On the other hand, it is *–tA*, if the last letter is Voiceless Consonant or Voiced Consonant of Type 2. Here *A* is for *a* or *e* according to last syllable of containing Front or Back Vowel.

In Table 4, boxes presented by numbers such as 1, 2 and 3 are used for personal possessive agreements in Table 2. For example, word *Eke*, "father" in Ablative case without a possessive agreement takes suffix *–den*, because the word *Eke* ends with vowel *e*. However, in third person possessive agreement it takes suffix *–nen*, because all words with third person possessive agreement in Ablative case always take suffix *–nen* even though the third person possessive agreement morpheme ends with vowel.

According to those similarities in Table 4, there are some generalized rules which are valid in many cases in grammar including verbs and derivations. Some of these generalized rules derived from close patterns given in Table 4, are given in Table 5. For example, Rule 12 in Table 5 represents rules for Locative and Dative cases in Group 1 in Table 4. In Table 4, Locative and Dative suffix rules are nearly identical and have same patterns which can be observed visually. Also, Accusative and Possessive Pronouns of Type 2 are same.

	1	2
1	**Rule 11** *Ablative Case*	**Rule 12** *Locative, Dative cases*
2	**Rule 21** *Genitive case*	**Rule 22** *Accusative case, Poss. Affix-2*
3	**Rule 31** *Plural Form of Noun*	**Rule 32** *Negation, Personal Agreement of A1Pl*
4	**Rule 41** *Instrumental case*	**Rule 42** *Personal Agreement of A1Sg*

Table 5. Generalized Rules.

In Dative case of GROUP 1 in Table 4, if the last letter is Back Vowel then *T* is replaced by *G* and *T* is replaced by *g* if the last letter is Front Vowel. Thus, a word *bala*, "child" becomes *bala-Ga*, "to child" and a word *Eke*, "father" will be *Eke-ge*, "to father". If the last letter is Voiceless Consonant, *T* is replaced by *q* or *k* depending on whether the last syllable contains Back Vowel or Front Vowel. For example, a word *kitap-qa*, "to book" has the last letter of Voiceless Consonant and the last syllable contains Back Vowel, thus *T* is replaced by *q*. A word *mektep-ke*, "to school" has the last letter of Voiceless Consonant and the last syllable contains Front vowel, thus *T* is replaced by *k*.

After detailed analysis of the language it can be seen that there are mainly common rules of alternations valid over all grammar. There are about 25 main alternation rules defined for all system together with generalized rules and 7 exception rules for each case. All these rules are implemented with XFST tools (Beesley and Karttunen, 2003). For instance, some mainly used common rules are given below and they are called by capital letters defined only in intermediate level. As mentioned before they are invisible by user. Here 0 is for empty character.

Rule H & Rule B: *H* is realized as 0 or *J*, *B* is realized as 0 or *A*.

```
[H->0,B->0||[Vowel]%+_[Cons]]
[H->J,B->A]
```

If the last letter of a morpheme is Vowel and the first letter of the following suffix is Consonant then *H* and *B* are realized as 0. Otherwise, they are realized as *J* and *B*. Some examples are:

ana-Hm→ *ana-m*, "my mother"
iS-Hm→ iS-Jm→Rule J→*iSim*, "my stomache"
ege-Br→ *ege-r*, "will sharpen"
bar-Br→ bar-Ar→ Rule A→*bar-ar*, "will go"

Rule J & Rule A: *J* is realized as *I* or *i* and *A* is realized as *y*, *a* or *e*.

```
[A->y||[Vowel]%+_]
[A->a,J->I||[BVowel](Cons)*%+?*_]
[A->e,J->i||[FVowel](Cons)*%+?*_]
```

If the last letter of a morpheme is Vowel then *A* is realized as *y*, and if the last syllable of a morpheme contains Back Vowel then *A* and *J* are realized as *a* and *I*. Otherwise, if the last syllable of a morpheme contains Front Vowel then *A* and *J* are realized as *e* and *i*. Some examples are:

bas-Hm→bas-Jm→*basIm*, "my head"
dos-tAr→*dos-tar*, "friends"
dEpter-lAr→ *dEpter-ler*, "copybooks"
barma-AmIn→*barma-ymIn*, "I will not go"

Rule T (a part of Rule 12 in Table 5)**:** *T* is realized as *q*, *G*, *k* or *g*.

```
[T->q||[BVowel](?)[VLCons]%+_]
[T->k||[FVowel](?)[VLCons]%+_]
[T->G||[BVowel](?)[0|SCons|VCons1]%+_]
[T->g||[FVowel](?)[0|SCons|VCons1]%+_]
```

This rule is a part of Rule 12 given in Table 5 for Dative case. It is one of generalized rules which are valid in many cases such as derivation of nouns, adjectives and verbs. Some examples are:

bala-Ta→ *bala-Ga*, "to child" (Noun in Dative)
Zaz-TI→ *Zaz-GI*, "of summer" (Adjective)
ZUr-Teli→*ZUr-geli*, "since coming" (Verb)
estit-Tiz→ estit-*kiz*, "make hear"(Causative Verb)

4 Nouns

Nouns in Kazakh Language take singular or plural (A3Sg, A3Pl) suffixes, Possessive suffixes, Case suffixes and Derivational suffixes. In addition, nouns can take Personal Agreement suffixes when they are derived into verbs. For example, *kitap-tar-da-GI-lar-dIN* which means "of those which is in books" has the following morphological analysis

kitap+Noun+A3Pl+Pnon+Loc^DB+Noun+Zero+A3Pl+Pnon+Gen.

Every nominal root at least has form of *Noun+A3Sg+Pnon+Nom*. Therefore, a root noun *kitap* which means "book" has the following morphological analysis

kitap+Noun+A3Sg+Pnon+Nom.

These inflections of noun are given in FST diagram in Figure 1.

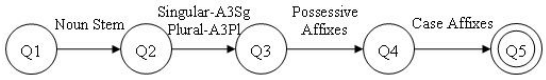

Figure 1. The FSA model of inflectional changes of a noun.

It can be seen that nominal root can be in singular form by adding (+0) no suffix which is in fact third personal singular agreement (A3Sg) and by adding suffix (+PAr) in plural form which is in fact third personal plural agreement (A3Pl). Here *P* is an intermediate level representation letter for *d*, *t* or *l* in surface level. After, possessive affixes (+Pnon:0, +P1Sg:Hm, +P2Sg:HN, +P2PSg:HNJz, +P3Sg:sJ, +P1Pl:HmJz, +P2Pl:HN, +P2PPl:HNJz, +P3Pl:s) and case affixes (Nom, Dat, Abl, Loc, Acc, Gen, Ins) are added. Here *H* and *J* are intermediate letters. All morphotactic rules together with adjective, pronoun, adverb and numerals are given in Figure 2. It can be observed that every adjective can be derived to noun and nouns with relative affix can be derived to adjectives. There are other derivations which are produced by adding some specific suffixes between verbs and nouns, adjectives and adverbs, adjectives and nouns. In order to get rid of complex view those derivations are not explicitly shown in Figure 2.

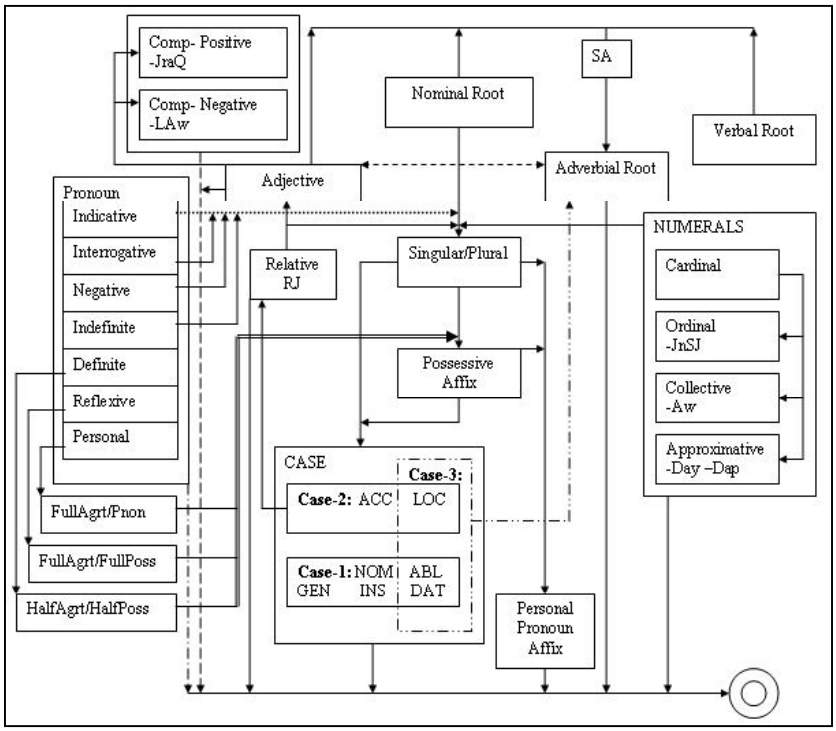

Figure 2. Morphotactic Rules for Nominal Roots.

In our system, the root of a word is a starting point for morphemes defined in lexicon file, and other morphemes are added according to morphotactic rules. Thus, starting from a root the system checks for all possible following morphemes and if a word is matched it gives appropriate output and moves to next state. For example, a surface form of a word *kitaptan,* "from a book" will have intermediate form of "kitap+tan" after implemented alternation rules. First it will check and find a noun root from lexicon. Then after giving output as "*kitap+Noun*", continues with next state which is *Singular/Plural*. At this state it will go on with 0 input giving output of *+A3Sg* for singular form of noun. Then, the next state will be *Possessive Affix* state to determine the personal possessive suffix. Here it is 0, thus epsilon transition which gives output as *+Pnon*. Now the output is *"kitap+Noun+A3Sg+Pnon"*. The next state is *Case* state in order to recognize the case of noun. Thus, for given input such as *+tan*, the output is determined as *+Abl* and this continues until the system reaches the final state or invalid state which is unacceptable state not returned to user. All possible morphemes are defined in the lexicon and all states are visualized in Figure 2.

	Saq/Tense	Suffix	
Future	bolZamdI keler Saq *Future Indefinite Tense*	ar/er/r	
	maqsattI keler Saq *Future Goal Oriented Tense*	baq/bek paq/pek maq/mek	
	auIspalI keler (osI) Saq *Future (Present) Transitional Tense*	a/e/y	Present
	maqsattI osI Saq *Present Goal Oriented Tense*	qalI/galI keli/geli	
	naq osI Saq *Present Definite Tense*	Ip/ip/p	
	dEl osI Saq *Present Progressive*	da/de	
Past	Zedel Otken Saq *Past Definite Tense*	dI/di tI/ti	
	auIspalI Otken Saq- Past *Transitional Tense*	atIn/etin ytIn/ytin	
	ayGaqtI burINGI Otken Saq *Past Narrative Definite Tense*	Gan/gen qan/ken	
	ayGaqsIz burINGI Otken Saq *Past Narrative Indefinite Tense*	Ip/ip/p	

Figure 3. Tenses of Verbs in Kazakh Language.

5 Verbs

Verbs are terms which define actions and states. Mainly three tenses exist such as present, future and past as stated in Figure 3. Moreover, conditional, optative and imperative moods are also defined. However in detailed form there are thirteen tenses together with modals in Kazakh language. These tenses are worked out from many resources where presentation and naming have variance among each other according to their scholars (Tuymebayev, 1996; Mamanov, 2007; Isaeva and Nurkina, 1996; Musaev, 2008). For example, according to Isaeva and Nurkina (1996) *awIspalI keler Saq* "Future Transitional Tense" denotes action in future and has same affix as Present Tense. However, Mamanov (2007) pointing out that *awIspalI keler Saq,* "Future Transitional Tense" denotes present action. Additionally, there are large amount of auxiliary verbs which define tenses and some modal verbs. However in cases that auxiliary verbs are not used verbs become as deverbial adverbs or participles which define verb or noun (Demirci, 2006). In Figure 4, morphotactic rules of verbs and modals are given. Derivations of verbs to nouns and adverbs with specific suffixes are shown with asterisk in Figure 4.

Verbs can be in reflexive, passive, collective and causative forms. For instance, verb *tara-w* means "to comb", *tara-n-w* in reflexive infinity form, *tara-l-w* in passive infinity form, *tara-s-w* in collective infinity and *tara-tQJz-w* and *tara-tTJr-w* in causative infinity form. Here, *Q*, *J* and *T* are intermediate letters. However not all verbs can have all of these forms at the same time.

Verbs in infinity form are generally formed with last letter w. For example: *kelw* which means "to come". The system is performing over generalization on verbs which take auxiliary verb on appropriate tenses. Those verbs are analyzed as derived adverbs or incomplete verbs on that tense since every verb of sentence should have personal agreement at the end and personal agreement affix added to the verb itself after the suffix of tense or to the auxiliary verb. In constructed morphological analyzer, we make analysis of every single word and for that reason generalization of some rules are made by giving more than one result.

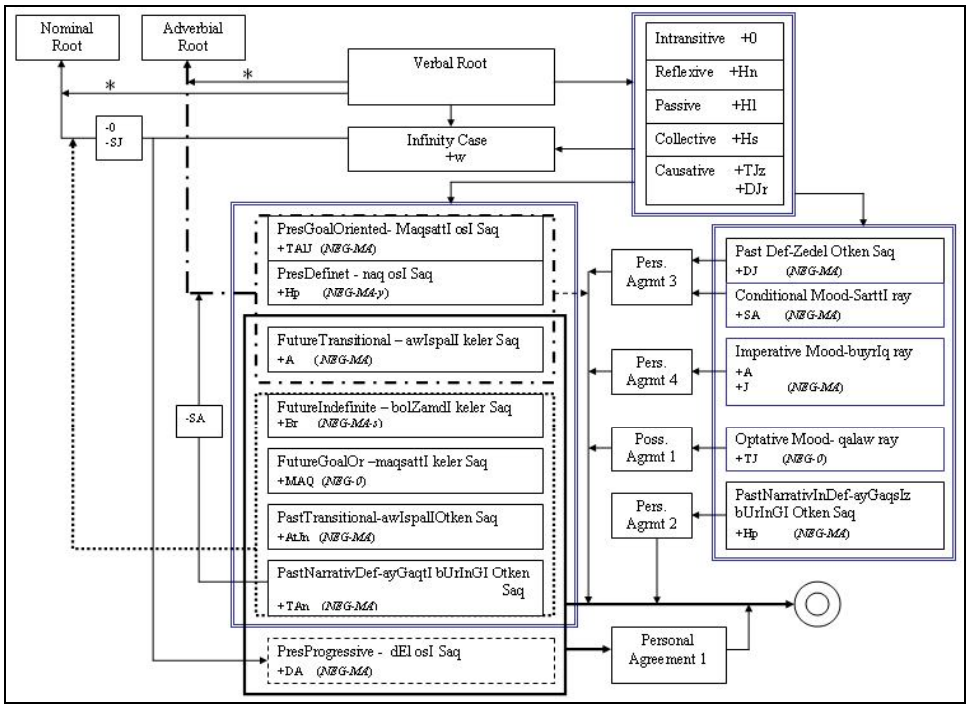

Figure 4. Morphotactic Rules of Verbs in Kazakh Language.

For example, *kel-geli tur-mIn* means "I am planning to come". Here *tur* is an auxiliary form which actually defines the tense of the verb and takes personal agreement affix *mIn*. Without an auxiliary verb, the word *kel-geli* means "since coming" and derived as an adverb. Thus compound verbs are examined separately. Some of tenses have different personal agreement endings and they are presented in Figure 4

6 Tests and Analysis

As mentioned before, the system is implemented using Xerox finite-state tools for NLP. Morphotactic rules and possible morphemes are defined in lexicon file and compiled with lexc tool. Alternation rules are defined in regex file and rule transducer is composed with lexicon file in one network with xfst tool. Loan words, proper names and technical terms are not included. System is working in two directions as in lexical and surface level. Due to the ambiguities in language there is no one-to-one mapping between surface and lexical forms of words and the system can produce more than one result.

A large corpus of Kazakh words (Qazinform, 2010) not seen by the morphological analyzer before has been continually analyzed in order to enhance the system by adding new words to lexicon. There are approximately 1000 words randomly selected from web which exist in lexicon and analyzed with the system. The percentage of correctly analyzed words is approximately 96%. Most of the errors are mainly the errors that occurred in the analysis of technical words which do not obey alternation rules of Kazakh Language. In Table 6, the w1.txt file has more technical words than w2.txt file. The results of the tests are given in Table 6. Errors due to Rules are exception errors which are not included in transducer yet. We hope in near future enhancing of the system will be performed by including all these rules. Also it can be seen in Table 6 that Kazakh words have 2.1 morphologic parses on average.

Files	Total Words	Correctly Analyzed Words	Total Errors	
			Rules	Analyzer
w1.txt	1000	962	30	8
w2.txt	1010	978	26	6
Morphologic Ambiguity is 2.1				

Table 6. Test Results.

7 Conclusion

Language is one of the main tools for communication. Thus its investigation provides

better perspectives on all other aspects related with NLP. However, formalization and computational analysis of Kazakh language morphology is not widely worked out. In other words, there is a lack of tools for analysis of Kazakh language morphology from computational point of view. Moreover, grammar resources contain variances depending on scholars. For example, in some resources there are twelve tenses, whereas in others there are much less tenses of verbs. Naming of tenses can also vary from source to source. To summarize, building correctly working system of morphological analysis by combining all information is valuable for further researches on language.

In this paper, a detailed morphological analysis of Kazakh language has been performed. Also, a formalization of rules over all morphotactics of Kazakh languages is worked out. By combining all gained information, a morphological analyzer is constructed. For future work, enhancing of system by adding exception rules related with loan words and proper names should be performed. Having more stabilized system with lessened possible rule errors some internal details of character encoding will also be solved. Moreover, releasing the working system to users on the web and collecting feedbacks are intended. These feedbacks from users can help on improving the system capacity and lessen any possible errors. This is planned to be performed with using an open source environment which is alternative to Xerox XFST, namely Foma by Hulden (2009).

Reference

Altenbek G and Wang X. 2010. Kazakh Segmentation System of Inflectional Affixes. *Proceedings of CIPS-SIGHAN Joint Conference on Chinese Language Processing* (CLP2010), Beijing, China, p.183–190.

Altintas K. and Cicekli I. 2001. A Morphological Analyser for Crimean Tatar. *Proceedings of the 10th Turkish Symposium on Artificial Intelligence and Neural Networks (TAINN'2001)*, North Cyprus, p.180-189.

Beesley R. K. and Karttunen L. 2003. *Finite State Morphology*. CSLI Publications, Stanford, CA, USA.

Coltekin C. 2010. A Freely Available Morphological Analyzer for Turkish. *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta.

Demirci K. 2006. *Kazakh Verbal Structures and Descriptive Verbs*. Dunwoody Press, Maryland, USA.

Isaeva S, Nurkina G. 1996. *Sopostavitelnaya tipologiya kazakhskogo i russkogo yazykov.* Uchebnogo Posobie. Sanat publishers, Almaty, Kazakhstan.

Hamzaoğlu I. 1993. *Machine translation from Turkish to other Turkic languages and an implementation for the Azeri language.* Master's thesis, Bogazici University, Turkey.

Hulden M. 2009. Foma: a finite-state compiler and library. P*roceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics: Demonstrations Session*. Association for Computational Linguistics, pp. 29–32.

Kairakbay M. B. and Zaurbekov D. L. 2013. Finite State Approach to the Kazakh Nominal Paradigm. *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing (FSMNLP 2013)*, Scotland.

Mamanov I.E. 1961. *Kazahskij jazyk*. Uchebnogo posobie gumanitarnogo tipa. Almaty, Kazakhstan.

Mamanov I.E. 2007. *Qazaq til biliminin maseleleri*. Aris publishers, Almaty, Kazakhstan

Mussayev M. K. 2008. *The Kazakh Language*. Vostochnaya literatura publishers, Moscow, Russia.

Oflazer K. 1994. Two-level Description of Turkish Morphology. *Literary and Linguistic Computing*, 9(2):137-148.

Orhun M., Tantuğ C., Adalı E., and Sönmez C. 2009. Computational comparison of the Uyg-hur and Turkish Grammar. *The 2nd IEEE International Conference on Computer Science and Information Technolog*y, pp:338-342, Beijing, China.

Qazinform. 2010. National news agency. http://www.inform.kz/qaz.

Salimzyanov I, Washington J. and Tyers F, 2013. A free/open-source Kazakh-Tatar machine translation system. *Machine Translation Summit XIV*, Nice, France.

Tantug C., Adali E. and Oflazer K. 2006. Computer Analysis of the Turkmen Language Morphology. *Lecture Notes in Computer Science*, 4139:186-193. A. C.

Tantug C., Adali E. and Oflazer K. 2007. AMT system from Turkmen to Turkish employing finite state and statistical methods. *Proceedings of MT Summit XI.*

Tuymebayev Q. Zhanseyit 1996. *Qazaq Tili: Grammatikaliq anaiqtagish*. Almaty, Kazakhstan.

Valyaeva T., 2014. *Kazakhskii yazyk.* http://www.kaz-tili.kz/.