Journal of Theoretical and Applied Information Technology
30th September 2016. Vol.91. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

ONTOLOGICAL MODELING OF MORPHOLOGICAL RULES FOR THE ADJECTIVES IN KAZAKH AND TURKISH LANGUAGES

[1] LENA ZHETKENBAY, [2]ALTYNBEK SHARIPBAY, [3]GULMIRA BEKMANOVA, [4]UNZILA KAMANUR

[1]Doctoral student, L.N. Gumilyov Eurasian National University, Astana, KAZAKHSTAN

[2] Prof., L.N. Gumilyov Eurasian National University, Astana, KAZAKHSTAN

[3]Assoc. Prof., L.N. Gumilyov Eurasian National University, Astana, KAZAKHSTAN

[4]Doctoral student, L.N. Gumilyov Eurasian National University, Astana, KAZAKHSTAN

E-mail: [1] jetlen_7@mail.ru, [2]sharalt@mail.ru, [3]gulmira-r@yandex.ru, [7]unzila.88@mail.ru,

ABSTRACT

This work aims to study the ontological model of morphological rules used in forming adjectives of Kazakh and Turkish languages. The analysis used in this work allows comparing the similarities and differences between the above-mentioned languages. The results taken can be used to create semantic translation systems that would be able to translate from Kazakh language into Turkish and vice versa, also for e-learning these languages through computers or the Internet.

Keywords: *Natural Language Processing, Adjective, Morphological Rules, Ontological Modeling, Machine Translation*

1. INTRODUCTION

The study of Natural Language Processing (NLP), which is reflected in this work, is a part of the computer science area "Artificial Intelligence". We consider the problem of computer processing of the Kazakh and Turkish languages, which are the part of the Turkic group of natural languages. It includes intelligent systems such as the machine translation from one language to another by using the automatic decision making and statement validation systems.

Kazakh language belongs to the Kipchak group [1], whereas Turkish language belongs to Oguz group of Turkic languages [2]. Both of the languages alike other Turkic languages have an agglutinative property, which is characterized by the ability of forming the word formation and word forms by adding affixes (suffixes) to the each root or stem of a particular word. In this case, suffixes change the semantics (meaning) of words which is part of the semantic category that forms new words, and ending - in the structural category, that can change only the composition of words. This property of Turkic languages is considered to be the main reason of ease of forming the morphological rules.

Nowadays, there are several methods used to form the morphological rules of the Turkic languages. Basically, they do not consider the semantics of words, and intended only for the formal description of the structure. In other words, it has been considered the rules that describe the suitable combinations of endings with the root of the word. There are only several studies related to Kazakh language which have been done by now [3], [4], [5]. The results of these studies have been used to develop systems that check the spelling of the Kazakh words or for e-learning purposes, but they do not make it possible to develop a machine translator based on the formal rules. All of that are due to the lack of semantic descriptions of words. For that reason, we aim to develop a system that would allow to develop the rules of word formation. In order to conduct the study we will use the ontological modeling method that allows semantically describe the grammar of natural languages. The research work includes introduction, related work, ontological modeling of morphological rules for the adjectives in Kazakh

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and Turkish languages, Conclusion and future work.

2. RELATED WORK

Currently, there are a variety of systems and methods of machine language translation [6], [7], [8]. The main reasons using these methods during the study are the difficulty of formalizing the language and the presence of the national languages corpuses. Studies considered in [9], [10] are the works on machine translation from Kazakh into other languages are based on grammatical rules. In these studies on the basis of Apertium system a bilingual database and structured database of grammatical rules (morphological, lexical, syntactic); algorithm and model lexical analysis; the technology of automatic generation of the grammatical rules of machine translation were developed and structured for the Kazakh-Russian and Kazakh-English.

The following papers were dedicated on the problems of machine translation from Turkish into other languages. In those works they performed by a finite-state two-level morphological analyzer for Turkmen Language and the same analyzer for the Uyghur Nouns by using Xerox Finite State Tools [11], [12]. The results of morphological analysis can be considered as an important step for language processing tasks of Turkic group, since no natural language processing system is designed without morphological analysis. In addition, a number of problems have been considered in order to improve the quality of bilingual corpuses of the statistical machine translation. Using the achievements of the abovementioned works significantly facilitates to the process of developing the machine translation from Kazakh into Turkish and vice versa.

3. ONTOLOGICAL MODELING OF MORPHOLOGICAL RULES FOR THE ADJECTIVES IN KAZAKH AND TURKISH LANGUAGES

Ontology is a powerful and widely used tool for modeling relationships between objects which belong to the different subject areas. It is possible to classify the ontology based on the degree of dependence on the task or application area, the ontological model of representation of knowledge and expression, as well as other criteria. The main part of the formal knowledge representations are based on the conceptualization: the objects, concepts, and other objects which exist in a certain area of interest and relationships defined between them. Conceptualization is an abstract, simplified view of the world that we want to present for some purposes [13]. In order to develop ontology first of all it is necessary to answer the following questions:

- What areas does an ontology cover? Answer: Adjectives.

- Why do we need an ontology? Answer: In order to develop a comparative ontological model for adjectives in Kazakh and Turkish languages.

- What kinds of questions should meet the information in the ontology? Answer: In order to conjugate by the meaning and the composition and define the types of adjectives.

- Who will use and maintain the ontology? Answer: linguists and programmers.

According to the above issues the ontological model of comparative adjectives will look like O (X, R, I), where X - names included in the structure of the adjectives (objects and concepts), R - the connection between names, I - a set of the name of these structures and relationships. The comparative ontological model of adjectives has been developed in the Protege environment (http://protege.stanford.edu). The Protege OWL language allows to describe not only the concept, but also the specific objects. Names and terms which are used ontological models in Kazakh and Turkish languages are given at table 1. The developed ontological model of adjectives of Kazakh language in Protege environment is displayed on Figure 1, and the ontological model of adjectives of Turkish language is shown in the Figure 2 below.

Thus, the comparative ontological model of adjectives designed for machine translation systems covers all components of the morphological analysis, such as: semantic groups of adjectives of Kazakh language are divided into relative and qualitative; morphological compositions are categorized as of individual and complex, based on form they are categorized as basic and derived. Basically, the adjectives according to the process of form are divided into morphological, syntactic, morphological and syntactic methods. They also can be converted according to the degree. The function Zhalghaw_zhalghanady was used in order to associate the adjectives with each other. The qualitative adjectives in Kazakh and Turkish are pluralizes (Koeptik), conjugates (Zhiktik) and inclines (Septik), more generally the suffixes and

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 1: Names And Terms Which Are Used Ontological Models In Kazakh And Turkish Languages

Kazakh	Turkish	English
Сын есім (Syn esim)	Sıfat	Adjective
Тұлғасына қарай (Tulghasyna qaraj)	*Yapılarına göre*	-
Негізгі (Negizgi)	Basit	Derivations
Туынды (Twyndy)	Türemiş	Derivatives
-	Birleşik	Compound
Құрамына қарай (Quramyna qaraj)	-	*Structure*
(Dara)	-	Simple
(Kuerdeli)	-	Complex
Сын есім жасайтын жұрнақтар (Syn esim zhasajtyn zhurnaqtar)	*Sıfatlar yapar yapım ekler*	*Adjective forming suffixes*
Есімдерден сын есім тудыратын жұрнақ (Esimderden syn esim twdyratyn zhurnaq)	Adlardan Yapılan Sıfatlar	Suffixes which form adjectives from noun, pronoun and numerals
Етістіктерден сын есім тудыратын жұрнақ (Etistikderden syn esim twdyratyn zhurnaq)	Eylemlerden Yapılan Sıfatlar	Suffixes which form adjectives from verbs
Қатыстық (Qatystyq)	Belirtme	Relative
Сапалық (Sapalyq)	Niteleme	Qualitative
Сын есімдердің жасалу тәсілдері (SE zhasalw taesilderi)	-	Adjectives are forming methods
Морфологиялық (Morphologyalyk)	-	Morphological
Синтаксистік (Sintaxsistik)	-	Syntactic
Морфологиялық- синтаксистік (Morphologyalyk-Sintaxsistik)	-	Morphological and syntactic
Жай шырай (Zhaj shyraj)	-	Absolute
Салыстырмалы шырай (Salystyrmaly shyraj)	Karşılaştırma	Comparative
Күшейтпелі шырай (Kueshejtpeli shyraj)	Pekiştirme	Superlative
Асырмалы шырай (Asyrmaly shyraj)		More superlative
Сапалық (Sapalyq)	*Niteleme*	Qualitative
-	Küçültme	Diminutive
-	Unvan*	Title
-	*Adlardan önce gelen	Titles before names
-	*Adlardan sonra gelen	Titles after names
-	*Hem adlardan once hem de sonra gelen	Titles before and after names
-	*Belirtme*	Relative
-	İşaret	Demonstrative
-	Soru	İnterrogative
-	Belgisiz	İndefinite
-	Sayı Sıfatlar*	Numeral adjectives
-	*Asıl Sayı Sıfatlar	Real number adjectives
-	*Sıra Sayı Sıfatlar	Ordinal adjective
-	*Üleştirme Sayı Sıfatlar	Distributive numeral adjective
-	*Kesir Sayı Sıfatlar	Numeral adjective
Жалғаулар (Zhalghawlar)	*Ekler*	*Endings*
Септік жалғауы (Septik	Hal durum ekleri	Case endings

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

zhalghawy)		
Көптік жалғауы (Koeptik zhalghawy)	Çoğul ekleri	Plural endings
Жіктік жалғауы (Zhiktik zhalghawy)	Şahıs eki	Personal endings
Тәуелдік жалғауы (Taweldik zhalghawy)	İyelik ekleri	Possesive endings

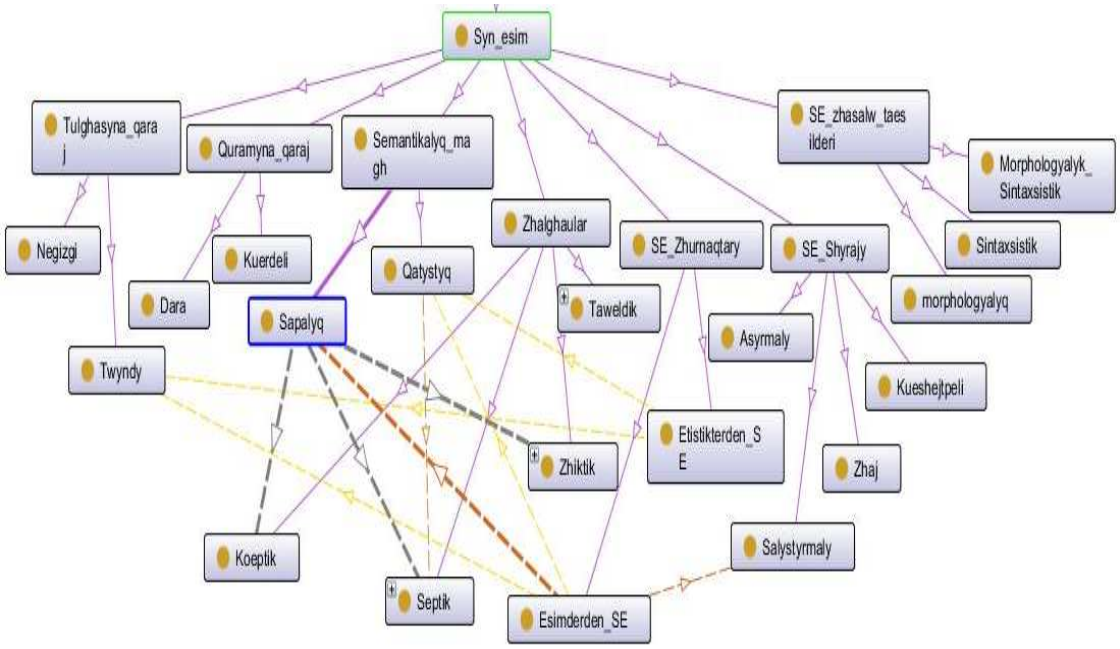

Figure 1: The Ontological Model Of Adjectives In Kazakh Language

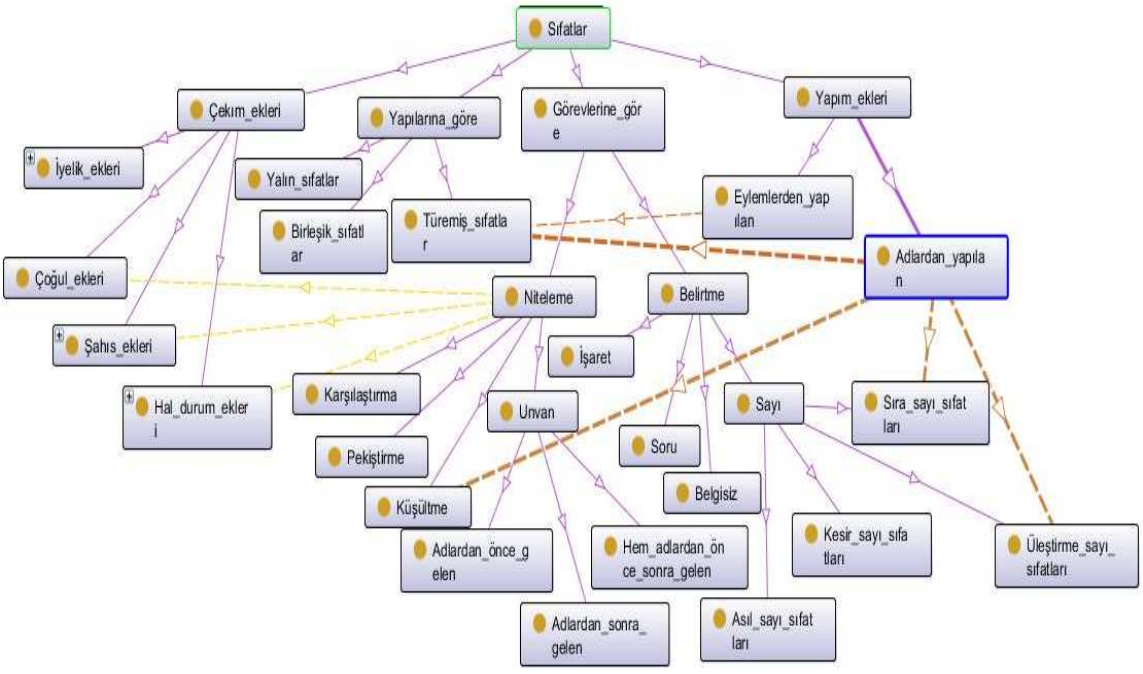

Figure 2: The Ontological Model Of Adjectives In Turkish Language

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

prefixes should be added. The relative adjectives only leans, thus adjectives substantiates. With the function Zhurnaq_arqyly_zhasalady it was shown that the comparative degree of adjectives is formed by adding a suffix to a word (name), and the relative adjective are formed not only from verbs, but also from the words (names) using suffixes. The adjectives in Turkish language are categorized as qualitative and relative (belirtme sıfatlar) adjectives. In terms of composition they can be basic, derivatives, compound (Birleşik). The qualitative (Niteleme) adjectives are divided into 4 types: comparative (Karşılaştırma), superlative (Pekiştirme), the diminutive (Küçültme) and rank (Unvan) [14]. The Turkish qualitative (Niteleme) adjective pluralizes (Koeptik), conjugates (Zhiktik) and inclines (Septik), so the adjective substantiates name and determinative adjectives (Belirtme sıfatlar) do not increases, conjugates and leans. By adding a suffix to the adjective derivative adjective, diminutive adjective (Küçültme), ordinal adjective, distributive numeral adjective (Üleştirme Sayı Sıfatlar), numeral adjective (Kesir Sayı Sıfatlar) are formed.

In the next point, we will consider the main differences and similarities between Kazakh and Turkish adjectives.

The superlative forms of adjectives in both languages are done by adding superlative suffixes to the main word (adjective). In Kazakh languages they are hyphenated, whereas in Turkish language they are written as one word. For example, in the Kazakh ap-achy, zhap-zhanga, zhip-zhingishke and Turkish yemyeşil, apacı, yepyeni, tertemiz, masmavi and etc.

In the next step, we will create conversion rules by which the superlative adjectives are formed in both Kazakh, Turkish languages and will consider formalizing using brackets method.

Note: First of all it is necessary to label letters of Kazakh and Turkish languages, we denote them as follows:

Kazakh letters	Turkish letters
AOUY!01	AOUI!01
AeOeUeIEJa IjJu!02	EİÖÜ!02
MNNg!03	MN!03
RWJ!04	RUY!04
L!05	L!05
BGGhD!06	BGĞD!06
ZhZ!07	CZ!07
P!08	P!08
!09	K!09
Q!10	H!10
STSh!11	STŞ!11

Next, we will use some labels to formalize the rules of adjectives of Kazakh and Turkish languages:

V_k – set of vowels in Kazakh language;

V_t – set of vowels in Turkish language;

C_k – set of consonants in Kazakh language;

A_k – set of adjectives in Kazakh language;

A_t – set of adjectives in Turkish language;

$\alpha, \beta, \gamma, \delta$ – variables which are formed by concatenation "$\cdot$" and length of their set are more than 1;

$\vee$ – the choice function (disjunction).

By using these labels we will write the rules of natural languages as masks in the form of $\frac{A}{B}$, where A- antecedents and B- consequents.

Rule 1 (for Kazakh language). If the main adjective in Kazakh language starts with a vowel then in the superlative forms of this adjective after the first vowel the letter ‘$\mathbf{p}$’ with a hyphen are added:

$$\frac{\delta \in A_k, \delta = \alpha \cdot \beta,\ \alpha \in V_k}{\alpha \cdot \mathbf{p} \cdot - \cdot \delta \in A_k};$$

Rules 1 (for Turkish language). If the main adjective in Turkish language starts with a vowel then in the superlative form of this adjective after the first vowel the consonant letter ‘$\mathbf{p}$’ is added:

$$\frac{\delta \in A_t, \delta = \alpha \cdot \beta,\ \alpha \in V_t}{\alpha \cdot \mathbf{p} \cdot \delta \in A_t};$$

Example 1, on the left side for Kazakh language, on the right side for Turkish language.

01X!((01p)-X)-ap-achy	01X!((01p)X)-apacı
01X!((01p)-X)-up-uzun	01X!((01p)X)-upuzun
01X! ((01p)-X)-yp-ystyq	01X!((01p)X)-ıpıslak
02X! ((02p)-X)-ep-eski	02X! ((02p)X)-epeski

Rules 2 (for Kazakh language). If the main adjective in Kazakh language starts with a consonant then in the superlative form of this adjective after the second letter the consonant letter ‘$\mathbf{p}$’ is added:

$$\frac{\delta \in A_k, \delta = \alpha \cdot \beta \cdot \gamma,\ \alpha \in C_k,\ \beta \in V_k}{\alpha \cdot \beta \cdot \mathbf{p} \cdot - \cdot \omega \in A_k};$$

Example 2, the superlative forms of adjectives in Kazakh language.

0301X!((0301p)-X)-mop-momyn	0901X!((0901p)-X)-kap-kara

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

0602X!((0602p)-X)-bip-bijik	1101X!((1101p)-X)-tap-taza
0601X!((0601p)-X)-dap-dajyn	1101X!((1101p)-X)-sap-sary

The superlative form of adjectives defines the level of difference of two objects, it points to the extent of advantages and disadvantages of an object. The Comparative adjectives in the Kazakh language are formed by using the following suffixes. In the Kazakh language: adjective +[raq, rek, yraq, irek law, lew, daw, dew, taw, tew, qyl, ghyl, qylt, ghylt, tym, shyl, shil, qay, ang]. In the Turkish language: daha (en) + adjective.

In Kazakh	In Turkish	In English
aqyldy-raq	daha akıllı	smart-er
zhyldam-yraq	daha hızlı	fast-er
arzan-daw	daha ucuz	cheap-er
kaeri-lew	daha yaşlı	old-er

4. CONCLUSION AND FUTURE WORK

The designed ontological models for computer processing of the Kazakh and Turkish languages in this work can be considered as an important step for comparative study of the two Turkic languages. For that reason, study of the structure and meanings of similar adjectives in Kazakh and Turkish languages, and the results of their comparison certainly gives a great opportunity for machine translation systems and the development of natural language processing systems.

As the result of the research work, we created an ontological model of the adjectives of Kazakh and Turkish languages and their rules are formed. On the basis of the rules, frequency of adjectives in Kazakh and Turkish language are created and automated.

It is believed that in the near future it will be developed an ontological model of morphological rules for the numerals, verbs and other parts of speech of the Kazakh and Turkish languages.

REFERENCES:

[1] Kazakh grammar, Phonetics, word formation, morphology, syntax (in Kazakh, Astana), 2002.
[2] Lewis, G., Turkish Grammar, Oxford University Press, 2001.
[3] Sharipbaev, A.A., "The basis of mathematical theory of Kazakh language" (in Kazakh), The collection of works of National Academy of Sciences of the Republic of Kazakhstan and the Institute of Informatics and management (Almaty), 1996, pp. 189-199.
[4] Sharipbaev, A.A., Bekmanova, G.T., "Formalization of morphological rules of Kazakh language with help of semantic neural network" (in Russian), Reports of the National Academy of Sciences of the Republic of Kazakhstan (Almaty), Vol. 4, 2009, pp. 11-16.
[5] Sharipbayev, A.A., Bekmanova, G.T., "Construction the logical semantics of the Kazakh words", Proceedings of the All-Russian conference with international participation "Knowledge-Ontologies-Theories" (Novosibirsk), Vol. 2, 2009, pp. 246-249.
[6] Koehn, F. J. Och, Marcu D., "Statistical phrase-based translation", Proceedings of NAACL-HLT (Edmonton, Canada), 2003, pp. 48–54,.
[7] Koehn, H., Hoang, A., Birch C., Callison-Burch, M., Federico, N., Bertoldi, B., Cowan, W., Shen, C.Moran R., Zens C., Dyer O., Bojar A., Constantin E. Moses, "Open Source Toolkit for Statistical Machine Translation", Proceedings of the ACL 2007 Demo and Poster Sessions, Prague, Association for Computational Linguistics, 2007, pp. 177–180.
[8] Lagarda, A.L., Alabau, V., Silva, F. R., D´ıaz-de-Lianono E., "Statistical Post-Editing of a Rule-Based Machine Translation System", Proceedings of NAACL HLT 2009: Short Papers, Boulder, Colorado, Association for Computational Linguistics, 2009, pp. 217–220.
[9] Abduali, B., Akhmadieva, Zh., Zholdybekova, S., Tukeyev, U., Rakhimova, D., "Study of the problem of creating structural transfer rules and lexical selection for the Kazakh-Russian machine translation system on Apertium platform", Proceedings of the International Conference Turkic Languages-2015, Academy of Sciences of the Republic of Tatarstan Press (Tatarstan), 2015, pp. 5-9.
[10] Tukeyev, U., Zhumanov, Zh., Rakhimova, D., Kartbayev, A., "Combinational Circuits Model of Kazakh and Russian Languages Morphology", Abstracts of International Conference "Computational and Informational Technologies in Science, Engineering and Education", Al-Farabi KazNU Press (Almaty), 2015, pp. 241-242.
[11] Tantuğ, A. C., Adalı, E., Oflazer, K., "Computer Analysis of the Turkmen Language Morphology", Lecture Notes in Computer Science, Springer, Vol. 4139, 2006, pp. 186-193.
[12] Orhun, M., Tantuğ, A. C., Adalı, E., "Rule Based Analysis of the Uyghur Nouns",

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Proceedings of the International Conference on Asian Language Processing (IALP), Chiang Mai, Thailand, 2008.

[13]Gruber, T.R.: "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", International Journal Human-Computer Studies, Vol. 43, 1995, pp.907-928.

[14]Hengirmen, M., Turkish Grammar (in Turkish) Ankara, 2007.