Semantic Neural Network Model of the Morphological Rules of the Kazakh and Turkish Noun for NLP System

Lena Zhetkenbay[1], Gulmira Bekmanova[2], and Altynbek Sharipbay[3]

Abstract **-** In this work there is illustrated an ontological model of nouns of Kazakh and Turkish languages and a comparison of the grammatical categories. On the basis of the ontological models, semantic neural networks of the morphological rules of nouns of the Kazakh and Turkish languages are developed and compared.

Keywords **-** agglutinative languages, morphology, ontology, semantic neural network.

I. Introduction

KAZAKH language belongs to Kipchak group [1] whereas Turkish language belongs to Oghuz group [2] of Turkic languages. The agglutination is one of the peculiarities of Turkic languages characterized by a large number of word types for each word formed by adding affixes to its end (suffixes and endings). There exists a strict order for adding of affixes: suffixes are the first to be attached to a stem or root of the word, then the endings of plural number, possessive endings, case endings and conjunction endings. All these features of Turkic languages assume a slight formalization for morphological and syntax rules as far as they have a strict order of adding affixes from a morphological point of view and a strict word order in sentence from the point of view of syntax. Therefore the issue of NLP can give decentish results on the basis of grammar rules.

At present the types of machine translation systems are varied. The choice in use of the machine translation system depends on the complexity of the formalization of natural language or national linguistic corpus of natural language.
The works [3-9] are devoted to the issues of machine translation from Turkish language into other Turkic languages. The issues of formalization of grammar rules in Kazakh language are solved in the works [10-20]. All these results will substantively facilitate the creation of Kazakh-Turkish machine translation system.

II. The Ontological Models of Nouns for Kazakh and Turkish Machine Translation System

Ontology is a powerful and widely used tool to model relationships between objects belonging to various subject fields. It is possible to classify ontologies based on the degree of dependence on the task or application area, the model of ontological knowledge representation and expressiveness, as well as other criteria [22, 23].

Applied ontologies describe concepts that depend on both the task and the subject domain of the ontology.

We used the ontology editor Protégé (http://protege.stanford.edu) to build the ontology. It is a free open source ontology editor and a framework for building knowledge bases. It was developed at Stanford University in collaboration with the University of Manchester.

Figure 1 illustrates the concepts and relationships which is used in the ontological models of nouns for Kazakh and Turkish machine translation system that used the notes in Protégé.

In this way, the comparisonal ontological models of noun for machine translation system includes all the categories of morphological features, for instance, noun is divided as base and complex according to structure of noun in Kazakh language, whereas in Turkish language there is not such division, furthermore, a noun can be common, proper, concrete, abstract, animated, inanimate according to meaning in Kazakh language, while in Turkish language a noun can be common, proper, animated, inanimated. In both languages the divisions of affixation are similar, e.g., the forms of cases, number, possessives and conjugations. There are seven cases whereas in Turkish there are five. The similarities of both languages are illustrated as basic notes in Table I.

Let's compare the similarities of cases and possessive forms of Kazakh and Turkish languages:

[1,2,3]L.N. Gumilyov Eurasian National University, Astana, 010008, Kazakhstan.

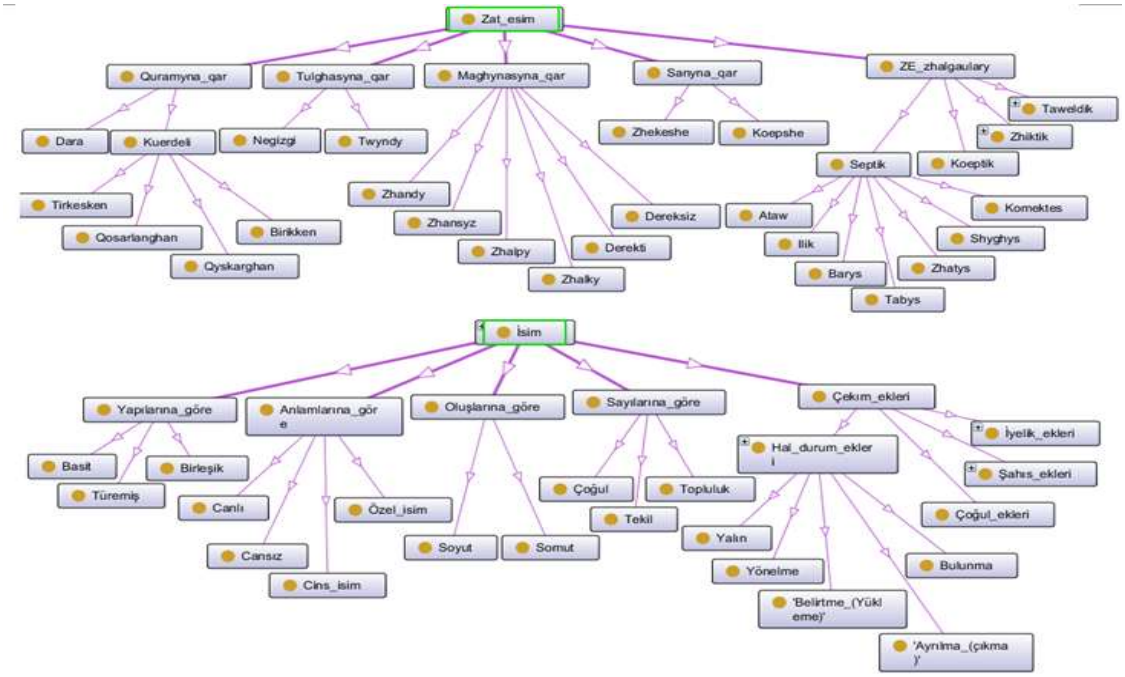

Fig. 1: The ontological models of nouns for Kazakh and Turkish machine translation system

TABLE I: COMPARE THE COMMON AND POSSESSIVE CASES OF KAZAKH AND TURKISH LANGUAGES

Cases		Common cases		Possessive cases	
Kazakh	Turkish	Kazakh	Turkish	Kazakh	Turkish
Atau septik (Nominative case)	Yalın hali	Үй (Uej)	Ev	Үйі(Uej-i)	Evi
İlik septik (Genitive case)		Үйдің(Uejding)			
Barys septik (Direction- dative case)	Yönelme hali	Үйге(Uejge)	Eve	Үйіне(Uejine)	Evine
Tabys septik (Accusative case)	Belirtme (Yükleme) hali	Үйді(Uejdi)	Evi	Үйін(Uejin)	Evini
Zhatys septik (Locative case)	Bulunma hali	Үйде(Uejde)	Evde	Үйінде(Uejinde)	Evinde
Shyghys septik (Ablative case)	Ayrılma (çıkma) hali	Үйден (Uejden)	Evden	Үйінен(Uejinen)	Evinden
Koemektes septik (Instrumental case)		Үймен(Uejmen)		Үйімен(Uejimen)	

As we see, there are seven cases in Kazakh language, while five in Turkish language.

III. SEMANTIC NEURAL NETWORK MODEL OF THE MORPHOLOGICAL RULES OF THE KAZAKH AND TURKISH NOUN FOR NLP SYSTEM

The example of formalization of rules of endings addition to stems is presented below. For formalization of the rules of endings and suffixes addition the semantic neural network presented in [24] is proposed for use. By means of such network Kazakh, Turkish languages word forms are generated and the structure of initial forms dictionary in the form of synchronized linear tree is produced.

For representation of a word form and its features the following metasymbols will be used:

- words separator,
(- word beginning,)
- word end,
! - word form feature beginning (case, etc.),
* - word form feature end.

Let us consider an example for Kazakh word "бала (bala) - child" and Turkish word "baba - father" (word stem) and its two word forms "балам (balam) - my child", "балаң (balang) - your child" (in the Kazakh, Turkish languages animate nouns change by persons by means of personal endings). The receptor is raised at word beginning symbol "(". Then it passes to "б" state, at supply of "б" symbol, then sequentially "(ба", "(бал", "(бала", and then simultaneously two substates "(балам)", "(ba", "(bab", "(baba", and then simultaneously two substates "(babam)" and "(балаң (balang))", "(baban)" (See Figure 2).

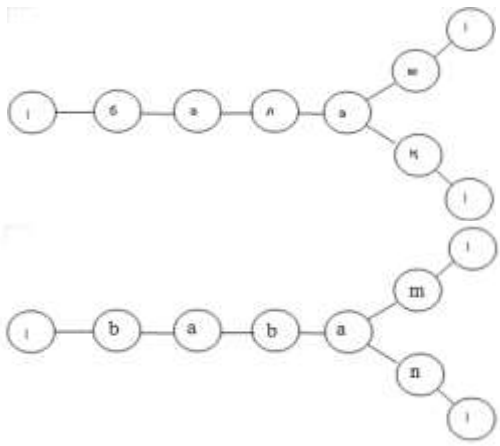

Fig. 2: Synchronized Liner Tree for the Kazakh, Turkish Words Forms with Endings Additions

Figure 3 shows an example of the structure of lemma connections defining the following features: noun (Зат есім) -

“зе*”, animate - “!жа*”, possessive ending (тәуелдік жалғау) of the first person - “!11*” (бірінші жақ). possessive ending (тәуелдік жалғау) of the second person - “!22*” (екінші жақ). At word “(балам)” supply to the lemma it passes to raised substates: “(балам) ”, “!зе*”, “!жа*”, “!11*” and at word "балаң" supply it passes to raised substates: “(балаң)”, “!зе*”, “!жа*”, “!22*”.

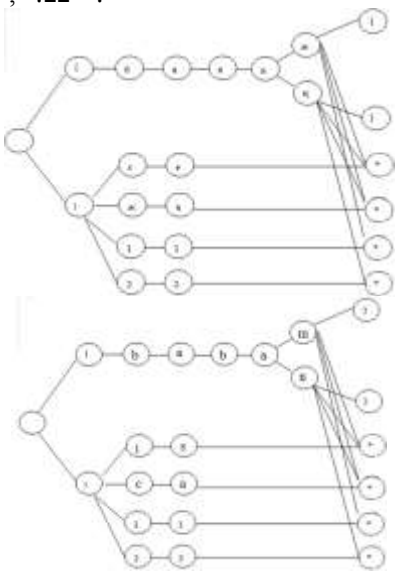

Fig. 3: Synchronized Liner Tree for Words Forms with Endings Addition and their Morphological information

Neurons-receptors distinguish separate symbols of input symbolical sequence. At the output the receptor generates a signal meaning either presence or absence of corresponding symbol in the analyzed text. Neurons-effectors provide the result of input symbolical sequence separate fragments recognition. In the synchronized linear tree we replace the signal from the receptor by the signal from the effector of the same tree. There will appear the opportunity of using symbolical sequences fragments as input symbols.

For indication of such fragments in the input symbolical sequence we will apply metasymbols of brackets: “(“ and Then the example presented will be rewritten in this form: ((бала)м), ((бала)ң), (((бала)м)ның), (((бала)ң)да), ((baba)m), ((baba)n), (((baba)n)da).

The fragment of formal inflexion rules by example of the noun subject to vowel harmony law conditioning soft or hard endings addition depending on the stem softness or hardness is presented below. The example below shows a fragment of rules where “зе”, “is” is noun (зат есім, isim), “жа”, “ca” is animateness (жанды, canlı), “01” ends by hard vowels a, o, ұ, (a, o, u) “))” between closing brackets endings of nouns are placed, after “!” there is morphological information.

Fragment of formal inflexion rules of the Kazakh, Turkish nouns

Rules (Kazakh)	Example	Rules (Turkish)	Example
((зежа01)мын)!жі11	((бала)мын)	((isca01)yım)!şe11	((baba)yım)
((зежа01)мыз)!жі11	((бала)мыз)	((isca01)yız)! çoşe11	((baba)yız)
((зежа01)сың)!жі22	((бала)сың)	((isca01)sın)!şe22	((baba)sın)
((зежа01)сыңдар)!жі22	((бала)сыңдар)		
((зежа01)сыздар)!жі22	((бала)сыздар)	((isca01)sız)! çoşe22	((baba)sız)
((зежа01))!жі33	((бала))	((isca01))!şe33	((baba))
((зежа01)м)!тә11	((бала)м)	((isca01)m)!iy11	((baba)m)
((зежа01)мыз)!тә1110	((бала)мыз)	((isca01)mız)!iy110	((baba)mız)
((зежа01)ң)!тә22	((бала)ң)	((isca01)n)!iy22	((baba)n)
((зежа01)ңыз)!тә22	((бала)ңыз)	((isca01)nız)!iy22	((baba)nız)
((зежа01)сы)!тә33	((бала)сы)	((isca01)sı)!iy33	((baba)sı)
((зежа01)лар)!кт	((бала)лар)	((isca01)lar)!ço	((baba)lar)
(((зежа01)лар)мыз)!ктжі11	(((бала)лар)мыз)		
(((зежа01)лар)сыңдар)!ктжі22	(((бала)лар)сыңдар)		
(((зежа01)лар))!ктжі33	(((бала)лар))	(((isca01)lar))!çoşe33	(((baba)lar))
(((зежа01)лар)ым)!кттә11	(((бала)лар)ым)	(((isca01)lar)ım)!çoiy11	(((baba)lar)ım)
(((зежа01)лар)ымыз)!кттә110	(((бала)лар)ымыз)	(((isca01)lar)ımız)!çoiy110	(((baba)lar)ımız)
(((зежа01)лар)ың)!кттә22	(((бала)лар)ың)	(((isca01)lar)ın)!çoiy22	(((baba)lar)ın)
(((зежа01)лар)ыңыз)!кттә22	(((бала)лар)ыңыз)	(((isca01)lar)ınız)!çoiy22	(((baba)lar)ınız)
(((зежа01)лар)ы)!кттә33	(((бала)лар)ы)	(((isca01)lar)ı)!çoiy33	(((baba)lar)ı)
((зежа01))!ат0	((бала))	((isca01))!ya0	((baba))
((зежа01)ның)!ілл	((бала)ның)		
((зежа01)ға)!ба	((бала)ға)	((isca01)ya)!yö	((baba)ya)
((зежа01)ны)!та	((бала)ны)	((isca01)yı)!be	((baba)yı)
((зежа01)да)!жс	((бала)да)	((isca01)da)!bu	((baba)da)
((зежа01)дан)!шы	((бала)дан)	((isca01)dan)!ay	((baba)dan)
((зежа01)мен)!кө	((бала)мен)		
((зежа01)менен)!кө	((бала)менен)		
(((зежа01)м)ның)!тә11ілл	(((бала)м)ның)		
(((зежа01)м)а)!тә11ба	(((бала)м)а)	(((isca01)m)a)!iy11yö	(((baba)m)a)
(((зежа01)м)ды)!тә11та	(((бала)м)ды)	(((isca01)m)ı)!iy11be	(((baba)m)ı)
(((зежа01)м)да)!тә11жс	(((бала)м)да)	(((isca01)m)da)!iy11bu	(((baba)m)da)
(((зежа01)м)нан)!тә11шы	(((бала)м)нан)	(((isca01)m)dan)!iy11ay	(((baba)m)dan)
(((зежа01)м)мен)!тә11кө	(((бала)м)мен)		
(((зежа01)м)менен)!тә11кө	(((бала)м)менен)		

The example of animate noun "бала" and "baba" inflexion contains all word forms of the given noun and their morphological information (MI) which contains in abridged notation the information on the number, case which the noun has, from what person the action takes place and its belonging to this or that person. Table II represents the inflexion of word example "бала" and "baba" by cases.

TABLE II: INFLEXION OF NOUN "БАЛА" (CHILD) AND "BABA" (FATHER)

Rules (Kazakh)	Explanation	Examples	Rules (Turkish)	Explanation	Examples
((зежа01))!ат*	(noun, animateness, 01 vowel)! nominative case	((бала))	((isca01))!ya*	(noun, animateness, 01 vowel)! nominative case	((baba))
((зежа01)ның)!ің*	((noun, animateness, 01 vowel)ның)! genitive case	((бала)ның)			
((зежа01)ға)!ба*	((noun, animateness, 01 vowel)ға)! dative case	((бала)ға)	((isca01)ya)!yö*	((noun, animateness, 01 vowel)ya)! dative case	((baba)ya)
((зежа01)ны)!та*	(noun, animateness, 01 vowel)ны)! accusative case	((бала)ны)	((isca01)yı)!be*	(noun, animateness, 01 vowel)yı)! accusative case	((baba)yı)
((зежа01)да)!жс *	((noun, animateness, 01 vowel)да)! locative case	((бала)да)	((isca01)da)!bu*	((noun, animateness, 01 vowel)da)! locative case	((baba)da)
((зежа01)дан)!шы*	((noun, animateness, 01 vowel)дан)! initial case	((бала)дан)	((isca01)dan)!ay*	((noun, animateness, 01 vowel)dan)! initial case	((baba)dan)
((зежа01)мен)!кө *	((noun, animateness, 01 vowel)мен)! instrumental case	((бала)мен)			

IV. CONCLUSION

Comparison of the ontological models of nouns of the Kazakh and Turkish languages allows to compare morphological rules for nouns of these languages. Comparison allows to formalize these rules by means of a semantic neural network for creation of the rule that based on the NLP systems.

REFERENCES

[1] Kazakh grammar. (2002). Phonetics, word formation, morphology, syntax (in Kazakh). Astana.

[2] Кононов А.Н. Грамматика современного турецкого литературного языка. – М-Л: Издательство АН СССР, 1956. - 570 с.

[3] Yıldırım E., Tantuğ A.C., "The feasibility analysis of re-ranking for N-best lists on English-Turkish machine translation", IEEE International Symposium on Innovations in Intelligent Systems and Applications, Albena, Bulgaria, 2013.

[4] Yıldırım E., Tantuğ A.C., Evaluation of Domain Adaptation Approaches to Improve The Translation Quality, 6th International Conference on Computational Collective Inte lligence Technologies and Applications (ICCCI 2014), Seoul, Korea, 2014.

[5] Tantuğ A.C., "Document Categorization with Modified Statistical Language Models for Agglutinative Languages", International Journal on Computational Intelligence Systems, vol. 5(3), 2010.

[6] M. Orhun, Tantuğ A. C., Adalı E., "Morphological Disambiguation Rules For Uyghur Language", IEEE International Conference on Software Engineering and Service Sciences (ICSESS), Beijing, China, 2010.

[7] Ilgen B., Adali E., Tantuğ A.C., A Comparative Study to Determine the Effective Window Size of Turkish Word Sense Disambiguation Systems, 28th International Symposium on Computer and Information Sciences, Paris, France, 2013.

[8] GulsenEryigit, JoakimNivre, and Kemal Oflazer. Dependency parsing of Turkish. Computational Linguistics, 2008. – pp.357–389.

[9] UmutSulubacak and GulsenEryi˘git. Representation of morphosyntactic units and coordination structures in the Turkish dependency treebank. In Proceedings of SPMRL 2013 (4th Workshop on Statistical Parsing of Morphologically Rich Languages), Seattle, USA, October, 2013.

[10] Бекманова Г.Т., Махимов А.К. Графематический анализ казахского неструктурированного текста // Информатизация общества: труды 3-ой международной научно-практической конференции.– Астана, 2012.– С. 509-511

[11] Бекманова Г.Т., Шарипбаев А. А. Формализация морфологических правил казахского языка с помощью семантической нейронной сети // Доклады Национальной академии наук Республики Казахстан. – 2009.– №4. – С. 11-16

[12] Bekmanova G., Sharipbaev A. The synthesis of word forms of Turkic language using semantic neural networks // Modern problems of applied mathematics and information technologies: abstracts – Al Khorezmy, ¬ 2009.–P.145.

[13] G.T.Bekmanova, A.A.Sharipbaeyv, S.K.Buribayeva Formalization of morphological rules of inflection in the Kazakh language // Вестник. Астана: Евразийский национальный университет им. Л.Н.Гумилева, 2012. – Специальный выпуск.– С.18-26

[14] A.Sharipbayev, G.Bekmanova, B.Yergesh, A.Buribayeva, and M.K.Karabalayeva. Intellectual morphological analyzer based on semantic networks. Processings of the OSTIS-2012, 2012, pp.397-400.

[15] Sharipbaev A.A. , Bekmanova G.T., Buribayeva A.K., Yergesh B.Z., Mukanova A.S., &Kaliyev A.K. (2012). Semantic neural network model of morphological rules of the agglutinative languages. In 6th International Conference on Soft Computing and Intelligent Systems and 13th International Symposium on Advanced Intelligence Systems, SCIS/ISIS, 1094–1099. http://dx.doi.org/10.1109/SCIS-ISIS.2012.6505337

[16] BanuYergesh, AsselMukanova, AltynbekSharipbay, GulmiraBekmanova, and BibigulRazakhova. Semantic Hyper-graph Based Representation of Nouns in the Kazakh Language. Computación y Sistemas Vol. 18, No. 3, 2014 pp. 627–635 ISSN 1405-5546 DOI: 10.13053/CyS-18-3-2041. http://dx.doi.org/10.13053/cys-18-3-2041

[17] Mukanova, A., Yergesh, B., Bekmanova, G., Razakhova, B., Sharipbay, A.Formal models of nouns in the Kazakh language. Leonardo Electronic Journal of Practices and Technologies.

[18] ШарипбаевА.А., Разахова Б.Ш. Formalization of syntactic rules of the Kazakh language // Специальныйвыпуск.- Астана: ЕНУим.Л.Н.Гумилева, 2012. - С.42-50;

[19] Sharipbayev A.A., Razakhova B. Sh. Mathematical models of syntactical rules of Kazakh language subject to semantics of parts of the sentence // The 4th Congress of the Turkic World Mathematical Society (TWMS). – Baku, 2011. –p. 463;

[20] L.Zhetkenbay, A.A.Sharipbay, G.T.Bekmanova, M.Khabylashimuly, U.Kamanur. The semantical, ontological models and formalization rules Kazakh compound words. Turklang'14 II International Conference on Computer processing of Turkic Languages, 2014.–Istanbul, 2014. – P.107-113. ISBN No.:978-975-561-449-6

[21] L.Zhetkenbay, A.A.Sharipbay, G.T.Bekmanova, U.Kamanur. The ontological model of noun for Kazakh-Turkish machine translation system. Turklang'15 III International Conference on Computer processing of Turkic Languages, 2015.–Russia, Kazan, Tatarstan, 2015. – P.15-24. ISBN No.:978-5-9690-0262-3

[22] Gruber T.R. A Translation Approach to Portable Ontology Specifications / Gruber T.R.// Knowledge Acquisition,1993, P.199-220

[23] Gruber T.R. Toward Principles for the Design of Ontologies Used for Knowledge Sharing / Gruber T.R. // International Journal Human-Computer Studies. – 1995, - Vol. 43 - P.907-928

[24] Shuklin, D.E. Structure of a semantic neural network that implements the morphological and syntactic analysis of text // Cybernetics and Systems Analysis. Kiev. Publishing House of the Institute of Cybernetics, 2001. - No 5. p. 172-179. In Russian