Semantic Hyper-graph Based Representation of Nouns in the Kazakh Language

Banu Yergesh, Assel Mukanova, Altynbek Sharipbay, Gulmira Bekmanova, and Bibigul Razakhova

L.N. Gumilyov Eurasian National University, Astana, Kazakhstan

b.yergesh@gmail.com, asel_ms@bk.ru, sharalt@mail.ru, gulmira-r@yandex.ru, utalina@mail.ru

Abstract. We explain how semantic hyper-graphs are used to describe ontological models of morphological rules of agglutinative languages, with the Kazakh language as a case study. The vertices of these graphs represent morphological features and the edges represent relationships between these features. Such modeling allows nearly one to one translation of the morphology of the language into object-oriented model of data. In addition, with such a model we can easily generate new word forms. The constructed model and the dictionary generated with it are freely available for research purposes.

Keywords. Ontology, morphology, semantic hyper-graph, word form generation.

1 Introduction

Modeling of morphology is relevant to all natural language processing applications and tasks [4] including information retrieval [7], sentiment analysis [35], spelling correction [8], detection of generated texts [6], part-of-speech tagging [23], and entity extraction [36]. Morphology is the study of the structure and formation of words. We discuss modeling of morphology of agglutinative languages, with the Kazakh language as a case study.

Agglutinative languages (lat. *agglutinare*: to glue together) are languages whose morphological system is characterized by agglutination ("gluing together") of different formants as the dominant type of inflection. The formants can be either prefixes or suffixes each having only one meaning.

The Kazakh language is a member of the Turkic group of languages; languages of this group can be classified as agglutinative languages. The Kazakh language has many word inflections. Inflections are formed by adding suffixes to words. Suffixes are attached in a strict sequence and resulting word forms vary in number, case, and person. The possessive form in Kazakh is similar to that in English [3, 19].

Graph-based models are widely used to represent natural languages [27].

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

An applied ontology is based on general principles of ontology building, and semantic hyper-graphs are used as a model for knowledge representation. This formalism determines ontology O as triple (V, R, K), where V is a set of concepts of a given subject field (hyper-graph vertex), R is a set of relationships between these concepts (hyper-graph edges), and K is a set of names of concepts and relationships in the domain.

The semantic hyper-graph language is a formal means of knowledge representation, in which it is possible to implement classifying, functional, situational, and structural networks and

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2041

scenarios, depending on the relationship types. This language is an extension of semantic networks, where N-ary relations are represented naturally; these relations not only allow us to specify the attributes of objects, but also permit representing their structural, "holistic" descriptions [21, 24].

Hyper-graphs generalize standard graphs by defining edges between multiple vertices instead of only two vertices [5]: a hyper-graph H = (V; E) on a finite set $V = \{v_i\}_{i \in I}$, I = {1, 2, ..., n}, of vertices is a family $E = \{e_j\}_{j \in J}$, J = {1, 2, ..., m}, of subsets of V called hyper-edges; I and J are finite sets of indexes.

A semantic hyper-graph is useful to model construction of morphological rules. It allows us to completely describe a morphological model by concepts and relationships.

In this paper, we use the semantic hyper-graph language to build a generative description of the Kazakh morphology, in particular, its noun subsystem. Both the model and the dictionary generated with it are freely available for research purposes.[1]

The paper is organized as follows. Section 2 provides a description of related work. Section 3 introduces our morphological model of the Kazakh noun. Section 4 presents the obtained results, Section 5 concludes the paper.

2 Related Work

There are many research works on modeling natural language morphology. The most widely used methods are finite state automata, two-level morphology, regular grammar, attribute grammar, and link grammar.

Finite state automaton is the most commonly used technique for implementing morphological analysis. It is applied for computational processing of many languages [16]. Automatic segmentation of inflectional affixes of the Kazakh language using finite state automata was studied in the work on a Kazakh segmentation system [2].

[1] http://e-zerde.kz/kazlang/ (mirrored at http://cys.cic.ipn.mx/data/18-3)

The authors of [17] presented finite state automata for the Kazakh nominal paradigm. This group of researchers developed a recommendation system for morphological analysis of Kazakh based on previously annotated data. For a given word or morpheme, the system generates a list of recommended markups ordered by the decreasing frequency of the previous usage [26]. Also for the Kazakh language, there exists an implementation of a generator as finite state automata [25].

Koskenniemi's two-level morphology [22] has been applied for modeling morphology of many languages, for example, Finnish [18], Turkish [28], Turkmen [34], Qazan-Tatar [14, 33], and Kazakh [37]. In [20], a rule-based morphological analyzer is developed using the two-level morphology approach. The construction of a rule-based morphological analyzer requires great effort.

Also for some inflective languages (Russian and Spanish), morphological analyzers were developed based on the method of analysis through generation [10, 11, 12], which avoids the use of complex analysis rules.

A general language-independent unsupervised algorithm for recognizing the morphological structure of an inflective language based on minimal length description was presented in [9]. This technique has been applied to English, Spanish, Russian, and a number of other languages. While the original algorithm only handled segmentation of words into two parts, it was extended to handle more morphs, which is more suitable for agglutinative languages [13].

Also, there have been suggested knowledge-poor algorithms for constructing word frequency lists with stemming which does not require any morphological dictionaries [1].

Works that discuss the use of semantic hyper-graph include [30, 38]. Zhen and Jiang [38] describe the semantic hyper-graph model as a hyper-graph based semantic network (Hy-SN), which can represent more complex semantic relationships and provides more efficient data structure for storing knowledge in repositories. However, the expressive power of all methods and approaches in modeling morphology is considered equivalent.

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2041

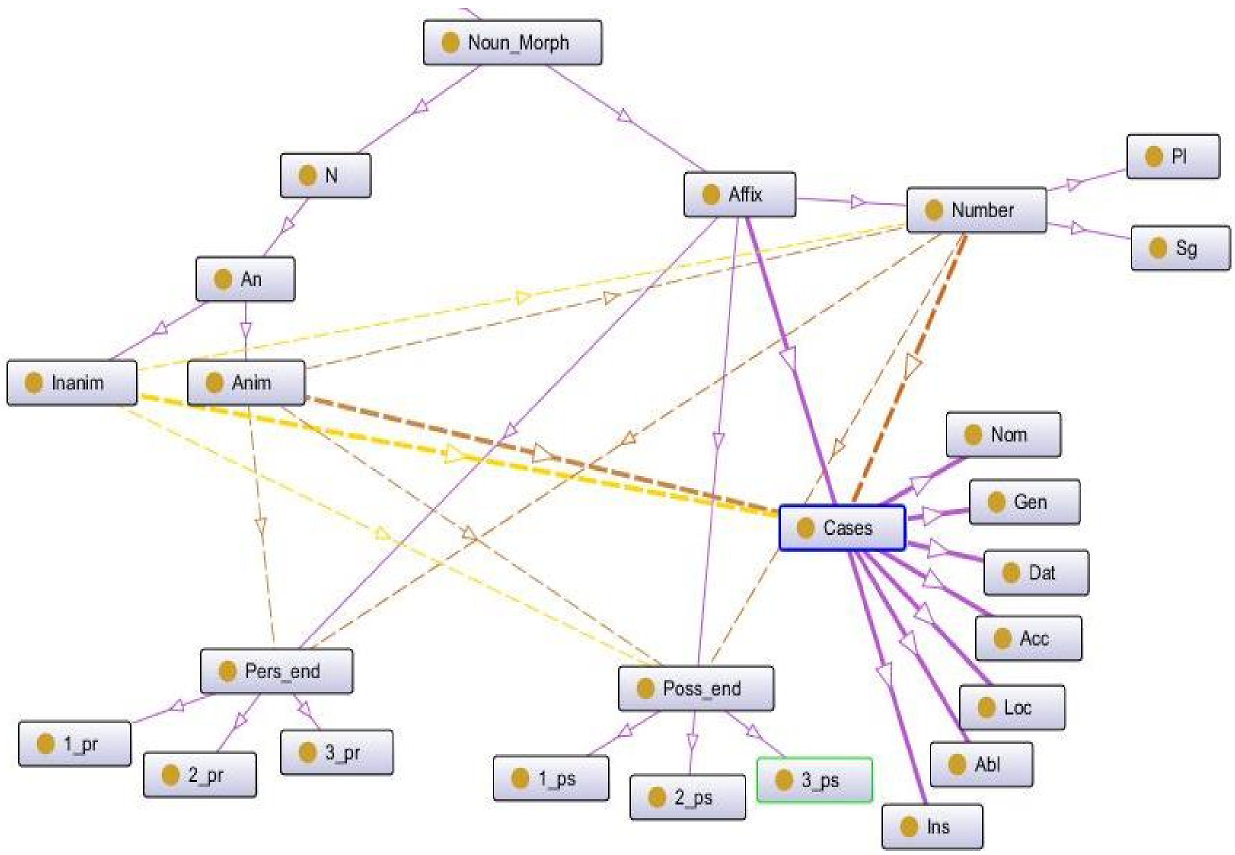

Fig. 1. Ontological model of the Kazakh noun

In this paper, our goal is to develop a simple implementation of the generator. Therefore, we propose a formal model described by using a semantic hyper-graph. Similarly to [32], the proposed model is easily implemented by object-oriented programming. Accordingly, it is easily modifiable, understandable, and readable. It provides implementation of high-quality morphological systems. The purpose of our research is automated generation of word forms and new words in the Kazakh language, as well as the morphological analysis of the Kazakh language.

Specifically, in this paper we give a detailed description of the noun subsystem. A model of verbal morphology will be presented in our future work.

3 Morphological Model of Kazakh Nouns

The morphological features of initial forms of nouns (N) are as follows. A noun can be either animate (anim) or inanimate (inanim); this feature determines the trajectory of the inflection of a noun. Nouns in the Kazakh language can be conjugated (pers_end) and vary for case (cases) and number (number), as well as have a possessive form (poss_end) [3, 19].

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

Table 1. Concepts and relationships

ID	Notation	Description
k_0	N	Noun
v_1	Part of speech	
v_2	Item	Item
k_1	Anim	Animate
v_3	Sign of animacy	
k_2	Inanim	Inanimate
v_4	Sign of inanimacy	
k_3	Cases	Cases
v_5	Nom	Nominative case
v_6	Gen	Genitive case
v_7	Dat	Direction- dative case
v_8	Acc	Accusative case
v_9	Loc	Locative case
v_{10}	Abl	Ablative case
v_{11}	Ins	Instrumental case
k_4	Pers_end	Personal endings
v_{12}	1 pr	1 personal

ID	Notation	Description
v_{13}	2 pr	2 personal
v_{14}	3 pr	3 personal
k_5	Poss_end	Possessive endings
v_{15}	1 ps	1 personal
v_{16}	2 ps	2 personal
v_{17}	3 ps	3 personal
k_6	Number	Number
v_{18}	Pl	Plural
v_{19}	Sg	Singular
e_1	is_a	
e_2	Denotes	
e_3, e_4	has_feature	
e_5, e_6	Has	
$e_7 – e_{21}$	Devided	
$e_{22} – e_{28}$	Change	
$e_{29} – e_{31}$	Add	

developed at Stanford University in collaboration with the University of Manchester.

Figure 1 shows the ontological model of the Kazakh noun with its morphological features. Concepts and relationships used in this ontological model are explained in Table 1.

The ontology model of the Kazakh parts of speech allows us to completely describe the morphological rules and their relationships. In addition, the representation of the ontology by means of a semantic hyper-graph allows translating the model into an object-oriented data model, where semantic hyper-graph vertices are classes. Figure 2 shows a graphical representation of the ontology using a semantic hyper-graph.

We will call hyper-edges semantic edges because they distinguish semantic hyper-graphs from other types of graphs; we will also assume that the set of vertices of the semantic hyper-graph includes a set of classes K = {k_a}, where a ∈ A = {0, 1, ..., n}, each of which consists of a set of instances of the class [29]. Thus, a vertex-class can be represented by a triple

$$k_a = \{\mathrm{V_a}, \mathrm{E_a}, \mathrm{S_a}\},$$

where $\mathrm{V_a}$ is a set of class properties, $\mathrm{E_a}$ is a set of semantic arcs incident to a given class, $\mathrm{S_a}$ is a set of instances of the class. The noun vertex-classes are

$$k_0 = \{\{\mathrm{v_1, v_2}\}, \{\mathrm{e_1, e_2}\}, \mathrm{S_1}\},$$
$$k_1 = \{\{\mathrm{v_3}\}, \{\mathrm{e_5}\}, \mathrm{S_2}\},$$
$$k_2 = \{\{\mathrm{v_4}\}, \{\mathrm{e_6}\}, \mathrm{S_3}\},$$
$$k_3 = \{\{\mathrm{v_5, v_6, v_7, v_8, v_9, v_{10}, v_{11}}\}, \{\mathrm{e_7, e_8, e_9, e_{10}, e_{11}, e_{12}, e_{13}}\}, \mathrm{S_4}\},$$
$$k_4 = \{\{\mathrm{v_{12}, v_{13}, v_{14}}\}, \{\mathrm{e_{14}, e_{15}, e_{16}}\}, \mathrm{S_5}\},$$
$$k_5 = \{\{\mathrm{v_{15}, v_{16}, v_{17}}\}, \{\mathrm{e_{17}, e_{18}, e_{19}}\}, \mathrm{S_6}\},$$
$$k_6 = \{\{\mathrm{v_{18}, v_{19}}\}, \{\mathrm{e_{20}, e_{21}}\}, \mathrm{S_7}\}.$$

We can represent the morphological noun model with the semantic hyper-graph model as a hyper-graph H(V,E), where

$$V = K = \{\mathrm{k_a}\}, \mathrm{E} = \{\mathrm{e_a}\},$$
$$V = \{\mathrm{k_0, k_1, k_2, k_3, k_4, k_5, k_6}\},$$

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2041

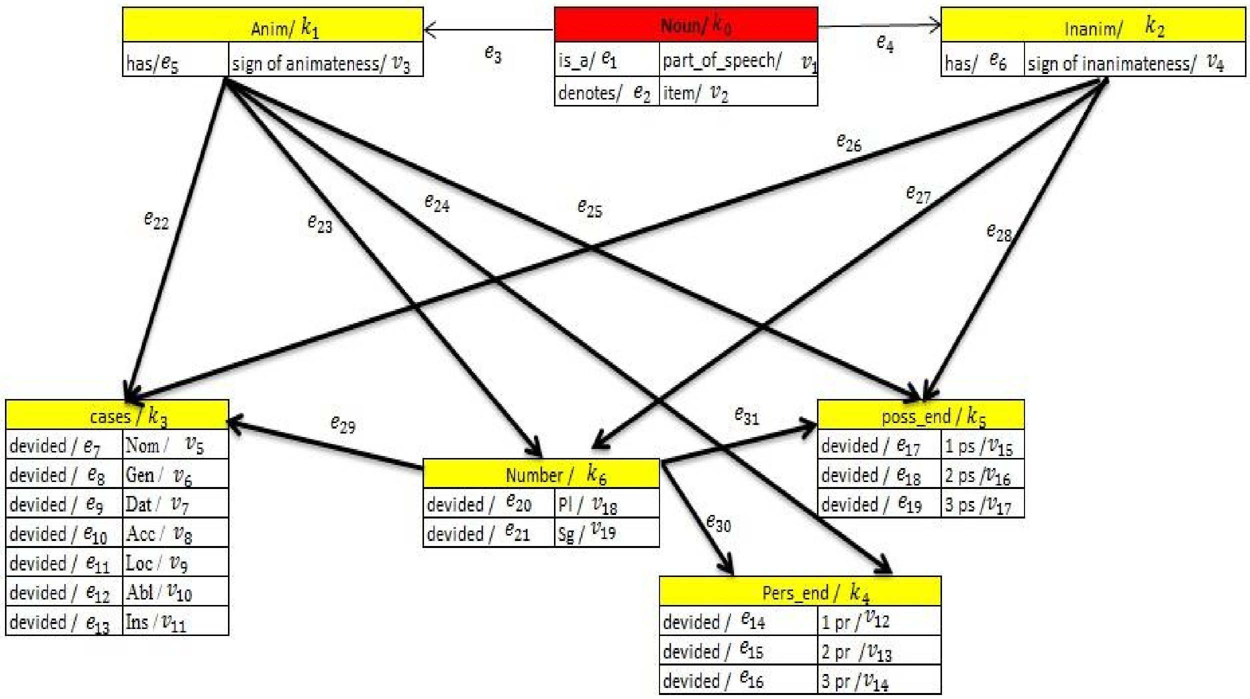

Fig. 2. Graphical representation of the ontology using a semantic hyper-graph

$$E = \{e_3 = \{k_0, k_1\}, e_4 = \{k_0, k_2\},$$
$$e_{22} = \{k_1, k_3\}, e_{23} = \{k_1, k_4\}, e_{24} = \{k_1, k_5\},$$
$$e_{25} = \{k_1, k_6\}, e_{26} = \{k_2, k_3\}, e_{27} = \{k_2, k_5\},$$
$$e_{28} = \{k_2, k_6\}, e_{29} = \{k_6, k_3\},$$
$$e_{30} = \{k_6, k_4\}, e_{31} = \{k_6, k_5\}\}.$$

4 Results

We constructed a database of initial forms which contains 40,000 words with their morphological features; 21,334 of these words are nouns. From the semantic hyper-graph described above, we obtained formal rules. The number of formal rules for nouns is 4,500.

Using these formal rules, 1,314,220 word forms of nouns were generated; it is also possible to generate nouns from stems of other parts of speech.

As an example, the inflection of the animate noun *bala* (child) includes all word forms of this noun and their morphological information in abbreviated notation which specifies the number, case, and person of a noun.

Example: inflection of the noun *bala*. The word *bala* is a noun. Noun (k_0) is a (e_1) part of speech (v_1), denotes (e_2) item (v_2), and answers the question “who?” It has a feature (e_3) of being animate (k_1). An animate noun in Kazakh changes (e_{22} – e_{25}) by cases (k_3), person (k_4), and number (k_6) and has a possessive form (k_5).

S = *bala*

Noun = {{Part_of_speech,Item}, {is_a,denotes}, {*bala* (child)}}

has_feature = {*bala*(child), anim},

change = {anim, cases}

cases = {{Nom, Gen, Dat, Acc, Loc, Abl, Ins}, {devided}, {*bala* (child), *balanyn* (child's), *balagha* (to the child), *balany* (child), *balada* (child), *baladan* (from child), *balamen*(with child)}}

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2041

change = {anim, pers_end}
Pers_end = {{1 pr, 2 pr, 3 pr}, {devided}, {{*balamyn* (I'm a child), *balasyn* (you are a child), *bala* (he/she is a child)}}

change = {anim, poss_end}
Poss_end = {{1 ps, 2 ps, 3 ps}, {devided}, {*balam* (my child), *balan* (your child), *balasy* (his/her child))}}

change = {anim, number}
Number = {{Pl, Sg}, {devided}, *bala* (child), *balalar* (children)}

Words in the Kazakh language can be inflected by adding some endings.

Add = {number, cases}
Add = {number, pers_end}
Add = {number, poss_end}

With the formal model of the noun described above, we can generate new word forms. In addition, this model can be used to model other parts of speech.

Based on these rules, we developed a generator of word forms for the Kazakh language. It can be used to create spellchecking technologies for the Kazakh language, translation applications, semantic search engines, speech recognition and speech generation technologies, etc.

Many methods of formalizing the morphological rules of a natural language do not allow any description of semantic properties of words. Our paper takes advantage of the possibility of using semantic hyper-graphs as a tool for formalizing the morphological rules of any natural language based on morphological features of words. Although this paper uses the Kazakh language to illustrate this concept, a semantic hyper-graph can be applied to any natural language.

The semantic hyper-graph representation of the morphological rules of the Kazakh language enables us to create formal rules of inflection and word formation for each part of speech. A software implementation of these rules made it possible to automatically generate more than 3.2 million word forms (dictionary entries) from 40,000 initial word forms with marked morphological features.

Earlier results were obtained using a semantic neural network. 2.8 million word forms were generated from 40,000 initial word forms [31]. The application of the semantic hyper-graph allowed us to increase the number of word forms by an order of magnitude, i.e., 400,000 units. This was achieved by a complete description of the morphological features of words using the expressive power of a semantic hyper-graph.

5 Conclusion

In this research, we obtained the following results:

- we created an ontological model of the noun of the Kazakh language;
- this ontological model is represented by a semantic hyper-graph;
- we obtained formal rules of generation of word forms;
- we generated a large dictionary of Kazakh word forms.

The resources developed by us are freely available for research purposes.[1]

References

1. **Alexandrov, M., Blanco, X., Gelbukh, A., & Makagonov, P. (2004).** Knowledge-poor Approach to Constructing Word Frequency Lists, with Examples from Romance Languages. *Procesamiento de Lenguaje Natural, No 33, 127–132.*
2. **Altenbek, G. (2010).** Kazakh Segmentation system of inflectional Affixes. In *Proceedings of CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP2010), Xiao-long Wang*. Beijing, China, 183–190.
3. **Batayeva, Z. (2012).** *Colloquial Kazakh*, Routledge.
4. **Bolshakov, I. A. & Gelbukh, A. (2004).** *Computational linguistics: Models, resources, applications*, 187 pp.
5. **Bretto, A. (2013).** *Hypergraph Theory*. Springer International Publishing Switzerland.
6. **Cinman, L., Dyachenko, P., Petrochenkov, V., & Timoshenko, S. (2013).** Automatic Distinction between Natural and Automatically Generated Texts Using Morphological and Syntactic

Information. *International Journal of Computational Linguistics and Applications*, Vol. 4, No. 2, 189–202.

7. **Gelbukh, A. (2000).** Lazy Query Enrichment: A Method for Indexing Large Specialized Document Bases with Morphology and Concept Hierarchy. *Lecture Notes in Computer Science*, Vol. 1873, 526–535.
8. **Gelbukh, A. (2003).** Exact and approximate prefix search under access locality requirements for morphological analysis and spelling correction. *Computación y Sistemas*, Vol. 6, No. 3, 167–182.
9. **Gelbukh, A., Alexandrov, M., & Han, S. (2004).** Detecting Inflection Patterns in Natural Language by Minimization of Morphological Model. *Lecture Notes in Computer Science*, Vol. 3287, 432–438.
10. **Gelbukh, A. & Sidorov, G. (2002).** Morphological Analysis of Inflective Languages through Generation. *Procesamiento de Lenguaje Natural,* No. 29, 105–112.
11. **Gelbukh, A. & Sidorov, G. (2003).** Approach to construction of automatic morphological analysis systems for inflective languages with little effort. In *Proc. of Computational Linguistics and Intelligent Text Processing, CICLing-2003, Mexico. Lecture Notes in Computer Science*, Vol. 2588, 215–220.
12. **Gelbukh, A. & Sidorov, G. (2005).** On Automatic Morphological Analysis of Inflective Languages. *International Conference on Applied Linguistics Dialogue-2005,* Russia, 92–96.
13. **Gelbukh, A., Sidorov, G., Lara-Reyes, D., & Chanona-Hernández, L. (2008).** Division of Spanish Words into Morphemes with a Genetic Algorithm. *Lecture Notes in Computer Science*, Vol. 5039, 19–26.
14. **Gökgöz, E. and Kurt, A. and Kulamshaev, K. & Kara, M. (2011)**. Two-Level Qazan Tatar Morphology. In *Proceedings of the 1st International Conference on Foreign Language Teaching and Applied Linguistics (FLTAL'11)*, 428–432.
15. **Gruber, T.R. (1995).** Toward Principles for the Design of Ontologies Used for Knowledge Sharing. *International Journal Human-Computer Studies* Vol. 43, Issues 5-6, 907–928.
16. **Jurafsky, D. & Martin, J. H. (2009).** *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. 2nd edition. Prentice-Hall.
17. **Kairakbay, B. & Zaurbekov, D. (2013).** Finite State Approach to the Kazakh Nominal Paradigm. In *Proceedings of the 11th International Conference on Finite State Methods and Natural Language Processing, 108–112.*
18. **Karttunen, L. (1983).** KIMMO: A general morphological processor. *Texas Linguistic Forum,* Vol. 22, 163–186.
19. **Kazakh grammar. (2002).** Phonetics, word formation, morphology, syntax (in Kazakh). Astana.
20. **Kessikbayeva, G. & Cicekli, I. (2014).** Rule Based Morphological Analyzer of Kazakh Language. In *Proceedings of the 2014 Joint Meeting of SIGMORPHON and SIGFSM,* Baltimore, Maryland USA, 46–54.
21. **Khakhalin, G. (2009).** Applied Ontology in the language of hypergraphs (in Russian). *Proceedings of 2nd All–Russian Conference "Knowledge—Ontology—Theory" (KONT-09).* Novosibirsk, 223–231.
22. **Koskenniemi, K. (1983).** *Two-level morphology: A general computational model for word-form recognition and production*. Publication 11, University of Helsinki.
23. **Lakshmana Pandian, S. & Geetha, T.V. (2008).** Morpheme based Language Model for Tamil Part-of-Speech Tagging. *Polibits*, Vol. 38, 19–26.
24. **Liu, H., LePendu, P., Jin, R., & Dou, D. (2011).** A Hypergraph-based Method for Discovering Semantically Associated Itemsets. In *ICDM 2011*, 398–406.
25. **Makazhanov, A., Makhambetov, O., Sabyrgaliyev, I., & Yessenbayev, Zh. (2014)** Spelling Correction for Kazakh. *Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science*, Vol. 8404, 533–541.
26. **Makhambetov, O., Makazhanov, A., Yessenbayev, Z., Matkarimov, B., Sabyrgaliyev, I., & Sharafudinov, A. (2013).** Assembling the Kazakh Language Corpus. In *EMNLP 2013*, 1022–1031.
27. **Martins, R. (2012).** Knowledge Vertices in XUNL. *Polibits*, Vol. 45, 61–66.
28. **Oflazer, K. (1994).** Two-Level Description of Turkish Morphology. *Literary and Linguistic Computing, Vol.* 9, No. 2, 137–148.
29. **Potchinskii, I. (2012).** Formal representation of semantic hypergraphs and their operations (in Russian). In *VIII International Scientific and Practical Internet Conference "Youth. Science. Innovation*s".
30. **Lian, R., Goertzel, B., Ke, S., O'Neill, J., Sadeghi, K., Shiu, S., Wang, D., Watkins, O., & Yu G. (2012).** Syntax-Semantic Mapping for General Intelligence: Language Comprehension as

ISSN 1405-5546
DOI: 10.13053/CyS-18-3-2041

Hypergraph Homomorphism, Language Generation as Constraint Satisfaction. Artificial General Intelligence. *Lecture Notes in Computer Science* Vol. 7716, 158–167.

31. **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.
32. **Sidorov, G. & Gelbukh, A. (1999).** A Hierarchy of Linguistic Programming Objects. In *Proc. ENC-99, 2º Encuentro de Computación*, Mexico, 12–15.
33. **Sulejmanov, D.Sh., Nevzorova, O.A, Gatiatullin, A.R., Gilmullin, R.A., Ayupov, M.M., & Pyatkin, N.V. (2007).** The main components of the application of the grammatical model Tatar (In Russian). In *Proc. Dialogue Conference 2007. Computational linguistics and intelligent technologies.* Russian State Humanitarian University, 525–530.
34. **Tantuğ, A. C., Adalı, E., & Oflazer, K. (2006).** Computer analysis of the Turkmen language morphology. *Advances in natural language processing. Lecture notes in Artificial intelligence,* Vol. 4139, 186–193.
35. **Wawer, A. (2012).** Extracting Emotive Patterns for Languages with Rich Morphology. *International Journal of Computational Linguistics and Applications*, Vol. 3, No. 1, 11–24.
36. **Zaraket, F. & Makhlouta, J. (2012).** Arabic Temporal Entity Extraction using Morphological Analysis. *International Journal of Computational Linguistics and Applications*, Vol. 3, No. 1, 121–136.
37. **Zafer, H. Reşit, Tilki, B., Kurt, A., & Kara, M. (2011).** Two-Level Description of Kazakh Morphology. In *Proceedings of the 1st International Conference on Foreign Language Teaching and Applied Linguistics (FLTAL'11),* 560–564.
38. **Zhen L. & Jiang Z. (2010).** Hy-SN: Hyper-graph based semantic network. *Knowledge-Based Systems*. Vol. 23, Issue 8, 809–816.

Banu Yergesh received her Master degree in Computer Science from the L.N. Gumilyov Eurasian National University of Astana, Kazakhstan, in 2011. She is currently a Ph.D. student at the L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2011, she has worked as a Lecturer at the Computer Science Department, Faculty of Information Technology of the L.N. Gumilyov Eurasian National University. At present she is a Researcher at the Scientific Research Institute "Artificial Intelligence." Her interests include artificial intelligence, computational linguistics, information retrieval, semantic analysis, models and methods for text classification.

Assel Mukanova received her Master degree in Computer Science from the L.N. Gumilyov Eurasian National University of Astana, Kazakhstan, in 2012. She is currently a Ph.D. student at the L.N. Gumilyov Eurasian National University, Kazakhstan. Since 2012, she has worked as a Lecturer at the Computer Science Department, Faculty of Information Technology of the L.N. Gumilyov Eurasian National University. At present she is a Researcher at the Scientific Research Institute "Artificial Intelligence." Her interests include artificial intelligence, computational linguistics, information retrieval, methods and algorithms for semantic knowledge processing, ontologies.

Altynbek Sharipbay is a Doctor of Technical Sciences, Professor of the specialty group "Computer Science, Computer Facilities and Management," a member of the International Informatization Academy, an academician of the Academy of Pedagogical Sciences of the Republic of Kazakhstan. He has authored 8 books, 2 monographs and 4 dictionary, published over 200 papers in conference proceeding and scientific journals and supervised several Bachelor, Master and Ph.D. theses. He participated in the development of many state standards. He is currently the Director of the Scientific Research Institute "Artificial Intelligence." He received state awards such as Winner of the State Prize of the Republic of Kazakhstan in the field of education, science and technology (2001), Medal for Merits in the Development of Science of the Republic of Kazakhstan (2003), Grant "The Best University Professor" of the Republic of Kazakhstan (2005, 2011). His interests include artificial intelligence, computational linguistics, speech recognition, software engineering, information security, and e-learning.

Gulmira Bekmanova received her Ph.D. degree in Computer Science, Computer Engineering and Control from the L.N. Gumilyov Eurasian National University of Astana, Kazakhstan, in 2010. She is a Candidate of Technical Sciences in Mathematics and Computer Software, Systems and Networks. She is currently the Head of the Computer Science Department, an Associate Professor of L.N. Gumilyov Eurasian National University. She has 45 scientific publications in scientific journals, national (Kazakhstan) and international conferences. Her interests include computer science, artificial intelligence, computational linguistics, speech recognition.

Bibigul Razakhova received her degree of Candidate of Technical Sciences in Mathematics and Computer Software, Systems and Networks from the L.N. Gumilyov Eurasian National University of Astana, Kazakhstan, in 2009. She is currently an Associate Professor of the Computer Science Department, Faculty of Information Technology of the L.N. Gumilyov Eurasian National University. She has 40 scientific publications in scientific journals, national (Kazakhstan) and international conferences. Her interests include computer science, artificial intelligence, computational linguistics, natural language processing.

Article received on 07/07/2014; accepted 10/08/2014.