Semantic Annotation of Verbs for the Tatar Corpus

Alfiya Galieva, Olga Nevzorova

Research Institute of Applied Semiotics of Tatarstan Academy of Sciences,
Kazan Federal University
e-mail: amgalieva@gmail.com, onevzoro@gmail.com

Abstract

This paper discusses the problem of developing the metalanguage for linguistic applications and introduces a tag set for the semantic annotation of verbs for the Tatar National Corpus. At present, there are no generally accepted standards for the development of corpus semantic annotation. In many cases, it is made by individual researchers or teams for one or another research project, and characteristics of tag sets used in thesauri and electronic corpora differ in many respects.

Using available semantic classifications of vocabulary for different languages and relying upon data from Tatar lexicons, we created a model of the semantic system of Tatar verbs and divided them into semantic classes (3,200 words). We distinguished semantic tags of two types: constructional (categorial) tags, independent of semantic classes of verbs, and semantic (thematic) tags, determining semantic classes of verbs. For the separation of these classes we used the hierarchical and overlapping classifications, so that the same verb may belong to more than one class. The approach is based on the data from explanatory dictionaries of the Tatar language, bilingual Russian-Tatar dictionaries and the system of semantic annotation of the Russian National Corpus. In the current version of our semantic annotation, we use 3 categorial and 59 thematic tags.

Keywords: Tatar verb; semantics; corpus; semantic annotation

1. Introduction

The development of the metalanguage for the semantic annotation of linguistic corpora is one of the topical issues in modern corpus linguistics. The system of annotation is a significant constituent of any corpus, and there are many kinds of corpus annotations: the morphological (PoS), syntactic, semantic, pragmatic and the anaphoric one. The extent of applicability and the potential of a corpus for achieving a variety of research and educational purposes depend in many respects on the thoroughness, accuracy and consistency of the annotation.

The semantic annotation for a corpus can be defined as additional information about the semantic category of words (Leech 2005). The semantic mark-up scheme embraces a set of semantic tags, the description of their meanings and the rules for assigning these tags to the units of a text or vocabulary. Usually, a morphological annotation distinguishing basic lexical and grammatical classes of words is used as a foundation for the semantic annotation of the vocabulary.

The paper discusses the problem of development of marking-up for linguistic applications and introduces a tag set for the semantic annotation of verbs for the Tatar National Corpus (http://corpus.antat.ru). The approach is based on the data from explanatory dictionaries of the Tatar language, bilingual Russian-Tatar dictionaries and the system of semantic annotation of the Russian National Corpus (http://www.ruscorpora.ru).

The Tatar National Corpus "Tugan Tel" ('Mother Tongue') was designed by a team of computational linguists from Research Institute of Applied Semiotics of Tatarstan Academy of Sciences (www.ips.antat.ru). At the moment, the corpus consists of written texts of different genres and styles of the modern literary Tatar language. The main sources of texts for the corpus are fictional texts, educational and scientific literature, texts of Internet publications on social and political topics and texts of official documents (Suleymanov et al. 2013). The corpus is grammatically annotated, and its estimated volume by the end of 2015 was 82 million words.

The paper is organized as follows: Section 2 is devoted to the background of the research. Section 3

represents available resources for classification of Tatar verbs. Section 4 discusses some semantic challenges that are significant for the project. Section 5 focuses on some aspects of the methodology and shows the types of the tags used in the research project.

2. Related Work

Verbs form a core of lexical and grammatical systems of any language; they have complicated semantic organization and require an integrated approach to the study. There are various semantic classifications of verbs made on vocabularies of different languages.

B. Levin, for instance, classifies English verbs (over 3,000 words) according to their similar meanings and syntactic behavior and separates the semantic classes taking into account a wide range of syntactic (mainly diathesis) alternations that reflect the verb meaning (Levin 1993).

The FrameNet project (https://framenet.icsi.berkeley.edu) implements the theoretical constructs of frame semantics (semantic frame, frame elements, frame-to-frame-relations, semantic types, etc.) in a lexical database of English basing on the annotation of examples of how words are used in actual texts (the meaning and usage of a verb). The project started at the International Computer Science Institute in Berkeley in 1997. FrameNet is based on the theory of Frame Semantics, deriving from the work of C. J. Fillmore and his colleagues (Fillmore 1976; Fillmore & Baker 2001; Fillmore et al. 2004). The project's key features are the commitment to corpus evidence for semantic and syntactic generalizations, and the representation of the valences of its target words in which the semantic portion makes use of frame semantic (Baker et al. 1998: 86).

The underlying principles of the English-based FrameNet project are successfully applied to the description and analysis of typologically diverse languages (Boas 2009). Currently FrameNet-like projects are developed for different languages. The paper of E. Kashkin and O. Lyashevskaya (2013) represents a research project in progress which involves a dictionary of Russian verbal lexical constructions and a corpus tagged with FrameNet-like annotation scheme. FrameBank project, originally intended as Russian analogue of Berkeley FrameNet, takes into account some recent approaches adopted in Construction Grammar and Russian lexical semantics, as well as certain features of the Russian lexical system and grammar. The mentioned paper focuses on the semantic annotation of constructions in FrameBank and describes the inventory of semantic roles used in FrameBank which correlates with the semantic classification of verbs and other predicates (Kashkin & Lyashevskaya 2013).

The Princeton WordNet (Miller 1995, Fellbaum 1998) and other wordnet-like thesauri (Vossen 2002) contain words of different parts of speech grouped into sets of synonyms (synsets), forming a hierarchical network. In the Princeton WordNet the verb synsets are arranged into hierarchies relying upon several entailment relations (Miller 1995, Fellbaum 1998), where prevalent is troponymy which maps the presence of the 'manner' relation between two verbs.

The VerbNet project (Kipper et al. 2006) is the largest computational verb lexicon currently available for English. It is organized into verb classes that are extended in comparison with those by Levin; each class is described by its thematic roles, selectional restrictions on the arguments, and syntactic frames. Another verb-oriented resource PropBank (Palmer et al 2005) represents a "proposition bank" of English verbs.

Lexical semantics is a long time developing issue of Russian lexicology. Semantic dictionaries of Russian verbs (Matveeva 1986, Babenko 1999) represent the results of semantic classifications of verbal lexis, where the classes are composed by shared meaning, and the basic deposition of classes is determined by such categories as action, state and relationship.

The National Corpus of the Russian Language (http://www.ruscorpora.ru) is provided by the system of annotation that enables search by lexical and semantic characteristics, based on a partial text mark-up. An extensive semantic classification of the lexicon is carried out separately for nouns, verbs, adjectives, adverbs and numerals. Developers appeal immediately to the meaning of the word and represent ontological (thematic) classification of words, and grammatical behaviour of a word is not taken into account (Kustova et al. 2009). This mark-up system attributes one or more

semantic and derivational tags to words; a multifaceted classification is used, so a word can fall into several classes. Since manual processing of the semantic mark-up of texts is very labour- and time-consuming, disambiguation is not carried out in the Corpus, and ambiguous words are attributed to several alternative sets of semantic features.

Basic semantic classes for the verbs of the Turkic languages are represented in special literature (Orasov 1983, Ganiev 1984); the researchers identify mainly 10-15 verb classes. Nevertheless, since there is no complete semantic classification of verbs accomplished on the basis of a large amount of vocabulary, there remain some questions related to the application of these classifications. The available corpora of the Turkic languages are not provided with any system of semantic annotation yet (the Tatar Corpus: http://corpus.antat.ru, the Crimean Tatar Corpus - http://korpus.juls.savba.sk/QIRIM/#id9, the Turkish Corpus - www.tnc.org.tr/index.php/en, the Kazakh Corpus - http://kazcorpus.kz/klcweb, the Bashkir Corpus - http://mfbl.ru/bashkorp/korpus, the Tuvan Corpus - http://www.tuvancorpus.ru, the Yakut Corpus - http://adictsakha.nsu.ru/corpora/corp).

At present the development of semantic annotations for a number of corpora of the Turkic languages is underway. The semantic tags in the electronic Khakass-Russian lexical database are set up within the framework of the project of the Corpus of Minority Turkic Languages; a preliminary version of the tags inventory is accomplished; unlike the semantic mark-up of the National Corpus of the Russian Language, the proposed system of semantic annotation takes into consideration not only paradigmatic, but also syntagmatic characteristics of the word meaning (Dybo et al., 2015).

The development of the semantic annotation of texts for the corpus of the Tuvan language is also started (Oorzhak & Khertek, 2015).

3. Available Resources

Currently we have no integrated description of the Tatar vocabulary as a complex hierarchical network of lexical units of different levels and types, and at present there are no special semantic dictionaries (thesauri) of the Tatar language yet, although the project of the Tatar wordnet-like thesaurus of verbs is in progress (Galieva et al., 2014, Galieva et al., 2015). This wordnet development takes into consideration such features of Turkic and Tatar verbs as co-existence of one-word and compound verbs, and a complicated system of grammatical voices.

This classification of Tatar verbs is made with the purpose of developing a corpus semantic dictionary, on the data of the available explanatory dictionaries of the Tatar language (1977-1981, 2005), bilingual Russian-Tatar dictionaries and the data from the Tatar National Corpus. When necessary, we also use the thesauri of the Russian language and the data from the Russian National Corpus; semantic classifications of other languages are also taken into consideration.

4. Challenges

The vocabulary of a language remains for researches the most complexly organized domain that is poorly amenable to systematization. Despite the well-known interest in semantics in the last decades, the generally accepted semantic metalanguage has not been created, and the task of constructing a metalanguage that would contain a finite, consistent, and foreseeable set of semantic features enabling one to describe the entire vocabulary, is far from being solved.

This can be explained by the existence of certain problems in the semantic description of a vocabulary:

- difficulty of differentiation of semantic components within a word,
- lack of distinct boundaries between taxons,
- necessity to take into consideration very large sets of characteristics and semantic components, and so on.

As a rule, semantic components do not have special formal indices (markers), so it is difficult, if not impossible, to distinguish them in a language and to describe them in a comprehensive and consistent way. Word meanings may be very rich, concrete and often unique in many respects, and

the issue of vocabulary systematizing outside certain classes (for example, kinship terms) remains open (Rakhilina & Plungian 2007).

At present, there are no standards for the development of corpus semantic annotation. In many cases it is made by individual researchers or teams for one or another research project. Such studies often involve non-standardized forms of annotating a data set, and the resulting annotated data are often not shared with other kinds of annotation that might be useful for specific research projects (Gries & Berez 2015). So quantitative and qualitative characteristics of sets of tags used in thesauri, electronic corpora and lexicographic databases differ in many respects.

Grammatical annotation uses a fixed set of grammatical classes. The number of semantic characteristics is uncertain and depends on the level of generalization. It is clear that the larger set of tags enables us the minute description of linguistic material. On the other hand, there are some advantages in simple encodings, for example, avoiding a great number of errors, eluding a huge amount of handwork, as well as consecution and consistency in the process of annotation. So, it is important to work out such an annotation system that would balance between specification and simplicity or transparency for developers' and users' convenience.

The verb is one of the most complicated parts of speech, capacious semantically and substantial grammatically. Verbs in the Turkic languages have an utterly ramified system of grammatical forms and meanings. Verbal semantics is not elementary, but comprehensive, which helps the verb represent discrete "pieces of reality", denoting elemental situations and events. So approaches to analysing verb semantics are based on different grounds and emphasize different layers of verb semantics.

The semantic classification of an entire array of words of a language does not only have applied, but also theoretical significance, for it may set the task of reviewing a number of theoretical propositions about organizing a vocabulary, linguistic denomination, building and structuring dictionaries. The subjects of numerous ongoing discussions are the methodology of analysis, the lists of semantic classes, the boundaries between the classes and attributing a verb to a particular class. The semantic classification of Tatar verbs also raises the issues of grammatical voice status, of the boundaries between the inflection and the word formation, as well as of polysemy and homonymy.

The main task in the course of our classification is to distinguish a semantic class of the verb. A reliable way for this is to refer the verb to a corresponding higher level concept (hypernim), for example, all motion verbs are semantically related to the verb *to move*. The work completed on semantic classification reveals that a significant part of verbal concepts of higher levels in Tatar are not verbalized; for instance, there are lacunas for hypernyms denoting concepts 'to have', 'to sound', 'to perceive', 'to create', etc.; nevertheless, low level concepts with rich and often complicated meanings for these semantic classes abound. So in many cases we are to gather individual words around artificially created higher level concepts.

Another challenge is anchoring or not anchoring lexical semantics to the derivational one. The agglutinative Tatar language has regular affixal ways of word formation, and some semantic subclasses are formed from verbs of the same derivational type. Example 1 represents Tatar instrumental verbs derived from nouns. The question is whether we are to distinguish such groups of words and give them individual tags or to combine them into larger classes.

(1) Example. Tatar instrumental verbs derived from nouns

Semantic class of verbs: Instrumental verbs
Tag for semantic class: *t:impact:tool*
Derivational model: noun +-*la* +(w)
Derivational meaning: 'to make a typical effect of using the tool'
Example of verbs: *borawlaw* 'to drill'; *tırmalaw* 'rake'; *carlaw* 'to sharpen'; *pıcaklaw* 'to cut using a knife'; *pıckılaw* 'to saw'; *kamcılaw* 'to lash'

Verb stem: *boraw* 'drill'; *tırma* 'to rake'; *car* 'wet stone'; *pıcak* 'knife'; *pıckı* 'saw'; *kamcı* 'lash'

The reason for this solution to assign such verbs special tags was due to the fact that these verbs have regular derivation and their list is open.

5. Methodology and Types of Tags

Usually a morphological annotation distinguishing basic lexical and grammatical classes of words is used as a foundation of the semantic annotation for vocabulary.

Taking into consideration our future task of unification of annotation systems of corpora for different languages and the perspective of building parallel corpora, we are aiming at developing a universal system of semantic annotation, where we rely upon existing approaches. Our tags are partially borrowed from the National Corpus of the Russian Language:

t:move (*sikerü* 'to jump');

t:ment (*fikerläw* 'to think', *aŋlaw* 'to understand');

t:perc — perception (*qaraw* 'to look', *isnäw* 'to smell');

t:speech (*söyläw* 'to talk', *lıgırdaw* 'to mumble');

t:sound (*qıştırdaw* 'to rustle', *žuıldaw* 'to buzz');

t:physiol (*yoklaw* 'to sleep', *awıru* 'to be sick'), etc.

In order to better reflect the singularity of semantics of Tatar verbs we use semantic tags of two types:

- constructional (categorial) tags, independent of the semantic class of the verb (Paducheva 2004: 45-46);
- semantic (thematic) tags, determining the semantic class of the verb.

If necessary, the auxiliary function of verbs is marked.

We distinguish the following categorial tags:

causative - *c:caus*: *qaytaru* 'to return', *taşlaw* 'to throw';

cooperative (reciprocal): - *c:coop*: *duslaşu* 'to become friends', *talaşu* 'to quarrel';

negative - *c:neg*: *tuqtaw* 'to stop', *tınu* 'to cease (about sound)'.

The tag *c:caus* is used for causative verbs; the tag *c:coop* is used for those verbs that require plurality of agents (a single agent is not permitted for the verb to be used); the tag *c:neg* is used for verbs that belong to a certain semantic class and is characterized by zero intensity.

The categorial components of the verb partly correspond with voice affixes; however they cannot be merely reduced to them, since in many verbs special voice affixes cannot be separated. For example, the Tatar language has causative verbs of two types:

a) morphological causatives are derivatives from corresponding non-causative verbs by means of causative affixes:

kaytu 'to return, to come back' - *kaytaru* (*kayt-ar-u*) 'to return sth';

basu 'to stand up, to rise' *bastıru* (*bas-tır-u*) 'to stand, to put sth';

b) lexical causatives do not contain any marker of causativity:

iltũ 'to take to, to carry sth.'

taŝlau 'to throw, to cast sth.'

All types of causative verbs are marked by means of the same special tag *c:caus*.

In the cases of significant differentiation of verbs within a semantic class we use basic and additional tags, which allows us to concretize the verb meaning. For example, the Tatar language has a great number of verbs denoting the beginning or becoming of a quality, and most of these verbs are derived from adjectives. They share the same inchoative derivational meaning; nevertheless their meanings are thematically different. Therefore we divide the verbs denoting quality change into several subclasses:

t:changest:humq (*qäbäxätlänü* 'to become vile'; *yawızlanu* 'to become malicious');

t:changest:color (*agaru* 'to become white', *qızaru* 'to become red');

t:changest:form (*turayu* 'to become straight', *yänçü* 'to flatten');

If a verb denotes a change of measured parameters we use tag *t:changest:param:*
awırayu 'grow heavy'.

The tag *t:changest:param* is used only for characteristics that may be expressed quantitatively.

In some cases the nature of a parameter is specified:

t:changest:param:cize (*zurayu* 'to increase in size', *qalınayu* 'become thicker');

t:changest:param:temper (*qaynarlanu*, *qızu* 'to become hot').

Since our classification is ontological (thematic) we may distinguish verbs related to different domains of reality. For example, for verbs denoting natural processes and phenomena we use special tag *t:nat:*

bayu 'to set (about sun)';

buranlaw 'to be a snowstorm',

isü 'to blow (about wind)',

yawu 'to fall (about rain or snow)'.

Figure 1 shows the use of tags designed for semantic annotation of the Tatar verbs.
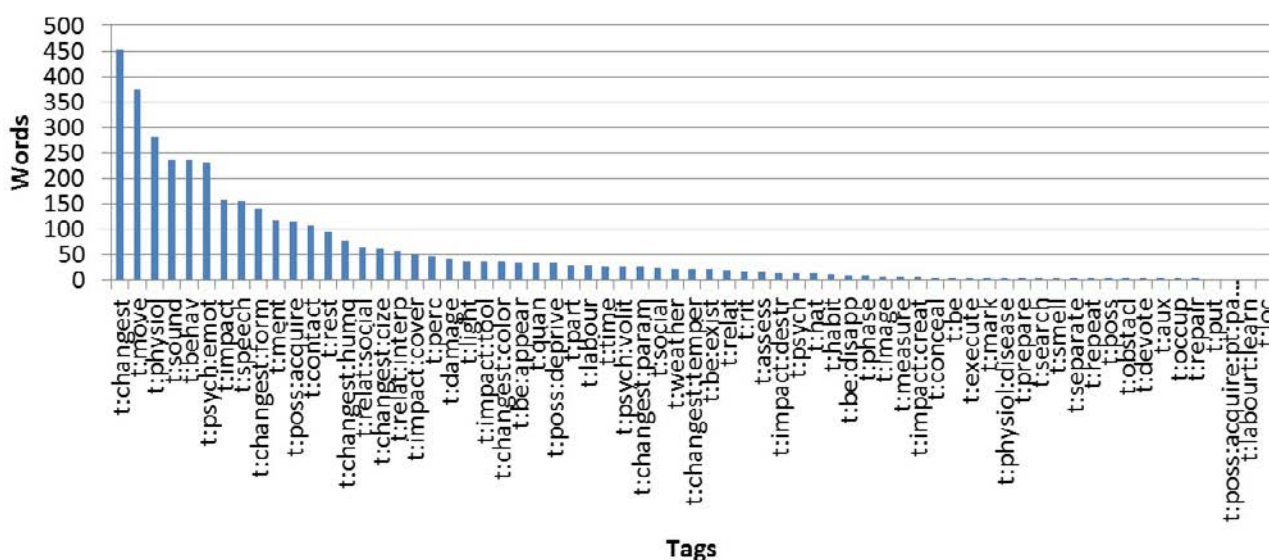

Fig1. Distribution of tags for semantic annotation of the Tatar verbs

A set of tags may be attributed to those verbs which are at the intersection of semantic classes (the overlapping classification principle is used):

kükräw 'to thunder (about thunder)' — *t:sound, t:nat;*

yäşnäw 'glitter (about lightning)' — *t:light, t:nat;*

ıŋgıraşu 'to moan' - *t:sound, t:physiol.*

Using a bunch of independent characteristics enables us to obtain a complex classification of a set of attributes.

A semantic feature is indicated by means of a special tag only if it occurs in the semantic structure of the fixed number of words. In the current version of annotation such threshold is 5 verbs.

6. Conclusion and Future Work

During the work on the corpus annotation for the Tatar language an inventory of semantic tags is being developed and the classification of verbs is being accomplished. For singling out classes we used the hierarchical and the overlapping classifications, so the same verb may belong to more than one class.

For verbs we distinguish between two types of tags: constructional (categorial) and semantic (thematic) tags. Constructional components of meaning are the same for all semantic classes and subclasses. Properly attributed semantic tags enables us to indicate classification headings.

The work on the development of the semantic annotation tag system for the Tatar National Corpus

is in progress, and 3,200 Tatar verbs have already been streamed into semantic classes. In the current version of the semantic annotation of verbs we use 3 categorial and 59 thematic tags (36 basic tags and 23 additional tags). The character and number of tags may be revised taking into consideration actual distribution of verbs in the Tatar Corpus. In the current version of annotation we rely exclusively upon thematic (ontological) classification. In future work grammatical behaviour of a verb, its participation in transformations of different kinds and its valence will be considered, which will allow us to complete and verify the classification performed.

The proposed system can be used for various linguistic applications for the Tatar language (text analysis systems, machine translation systems).

References

Babenko L. (ed.). (1999). *Explanatory Dictionary of Russian Verbs: an Ideographic Description.* Moscow.

Baker, C. F., Fillmore, C. J. & Lowe, J. B. (1998, August). The Berkeley FrameNet project. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1* (pp. 86-90). Association for Computational Linguistics. 86-90.

Boas, H. C. (Ed.). (2009). *Multilingual FrameNets in computational lexicography: methods and applications* (Vol. 200). Walter de Gruyter.

Dybo A., Sheymovich A. & Krylov S. (2015). Some Possibilities of Semantic and Etymological Tagging of Corpora for Turkic Languages. In *Proceedings of the International Conference "Turkic Languages Processing" TurkLang-2015,* Kazan, pp. 304-327.

Fellbaum, C. (1998). *WordNet. An Electronic Lexical Database*. Cambridge, Mass: MIT Press.

Fillmore, C. J. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences*, *280* (1). pp. 20-32.

Fillmore C.J. & Baker, C. F. (2001). Frame Semantics for Text Understanding. In *Proceedings of WordNet and Other Lexical Resources Workshop*, Pittsburgh. Accessed at: http://www.ccs.neu.edu/course/csg224/resources/framenet/framenet.pdf [30/04/2016].

Fillmore C.J., Baker, C. F. & Sato, H., FrameNet as a "Net" In *Proceedings of LREC*, Lisbon, 2004, vol. 4, pp. 1091-1094.

FrameNet. Accessed at: http://framenet.icsi.berkeley.edu [19/04/2016]

Galieva A, Nevzorova O.& Gatiatullin A. (2014). Towards Building Wordnet for the Tatar Language In *Communications in Computer and information Science,* Vol. 468. Knowledge Engineering and the Semantic Web. 5th International Conference on Knowledge Engineering and Semantic Web. September 29 – October 1. Proceedings, Springer, pp. 57-66.

Galieva A., Nevzorova O. & Suleymanov D. (2015). Corpus Based Tatar Lexicography: Verbs in TatWordnet. In *Procedia - Social and Behavioral Sciences,* 198 (2015), pp. 132 – 139.

Ganiev, F.A. (1984). Semantic Classes of Tatar Verbs. In *Studies on Tatar linguistics*. Kazan, pp. 75-84. (In Russian).

Gildea D., Jurafsky D. (2000). Automatic Labeling of Semantic Roles. In *Proceedings of the 38th Annual Conference of the Association for Computational Linguistics (ACL '00)*, pp. 512–520.

Gries S. & Berez A. (2015). Linguistic Annotation in/for Corpus Linguistics. In *Handbook of Linguistic Annotation*. Berlin, New York: Springer.

Kashkin E. V., Lyashevskaya O. N. (2013). Semantic roles and construction net in Russian FrameBank In Computational Linguistics and Intellectual Technologies. Procedings of "Dialogue" International Conference . Volume 12, Moscow, RSUH. 297-311.3, pp/ 325-343 (In Russian).

Kipper, K., Korhonen A., Ryant N. & Palmer M. (2006). Extending VerbNet with Novel Verb Classes. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation -- LREC'06*. May, 2006, Genoa, Italy: 2006.

Kustova G., Lashevskaja O., Paducheva E.& Rakhilina E. (2009), Verb Taxonomy: from Theoretical Lexical Semantics to Practice of Corpus Tagging. In Lewandowska B., K. Dziwirek

(eds.), *Cognitive Corpus Linguistics Studies*. Frankfurt: Peter Lang.

Leech, G. (2005). Adding Linguistic Annotation, Chapter 3. In M. Wynne (ed.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, pp. 17-29.

Levin B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation.* Chicago, University of Chicago, 1993.

Lyashevskaya O. (2010), Bank of Russian Constructions and Valencies // Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), Valletta, Malta, 17–23 May 2010. Valletta: ELRA, 2010, pp. 1802–1805.

Matveeva T. (ed.) (1986). *Lexical Semantic Groups of Russian Verbs (an Academic Dictionary).* Sverdlovsk. (In Russian).

Miller, G. A. (1995). WordNet: A Lexical Database for English. In *Communications of the ACM* Vol. 38, No. 11: 39-41.

National Corpus of the Russian Language. Accessed at: http://www.ruscorpora.ru [29/04/2016].

Oorzhak B., Khertek A. (2015). Development of Semantic Markup for the Corpus of Tuvan Language. In *Proceedings of the International Conference"Turkic Languages Processing" TurkLang-2015*, Kazan. Pp. 351 – 373

Orazov M. (1983). *Semantics of the Kazakh Verb (an Experience of Semantic Classification).* Alma-Ata. (In Russian).

Paducheva E. (2004). *Dynamic Models in Semantics of Vocabulary.* Moscow. (In Russian).

Palmer M, Kingsbury P, Gildea D (2005). "The Proposition Bank: An Annotated Corpus of Semantic Roles". Computational Linguistics 31 (1): 71–106.

Rakhilina E. & Plungian V. (2007). *About Lexical Semantic typology*. In *Verbs of Movement in Water: Lexical Typology.* Moscow, pp. 11-26. (In Russian).

Suleymanov, D., Nevzorova, O., Gatiatullin A., Gilmullin, R. & Khakimov, B. (2013). National Corpus of the Tatar Language "Tugan Tel": Grammatical Annotation and Implementation. In *Procedia - Social and Behavioral Sciences,* 2013, pp. 68-74.

Tatar National Corpus. Accessed at: http://corpus.antat.ru [29/04/2016].

The Tatar Explanatory Dictionary in 3 volumes (1977-1981), Kazan (In Tatar).

*The Tatar Explanatory Dictionary in 1 volum*e (2005). Kazan. (In Tatar).

Vossen P. (ed.) (2002). *EuroWordNet General Document. Version 3.* Accessed at: http://vossen.info/docs/2002/EWNGeneral.pdf [29/04/2016].