THE SEMANTIC MODEL OF A LINGUISTIC KNOWLEDGE BASE

Olga Yeliseyeva, Yury Kim

Abstract: *This paper suggests basic principles for creating the semantic model of a linguistic knowledge base (SMLKB). The approach we're using here allows us to unify all stages of the automated processing of natural language structures. The given model enables us to create applied natural language systems using the same principles as with knowledge-based systems. The new SMKLB also makes possible deep-level language research and the creation of corresponding knowledge bases to store the results of such research. For the developers of applied intelligent systems SMKLB can become a starting point for a more efficient creation of natural language interfaces. Finally, language intelligent tutoring systems could be created within the new approach.*

Keywords*: knowledge base, linguistic knowledge base, semantic model.*

ACM Classification Keywords*: I.2 ARTIFICIAL INTELLIGENCE - I.2.4 Knowledge Representation Formalisms and Methods - Semantic networks, I.2.7 Natural Language Processing - Language models.*

Introduction

When formalizing a language one must distinguish the different levels and sub-levels of its structure (at the most general level, morphology, syntax and semantics ought to be represented) as well as describe the rules of transition between these levels in order to ensure the functioning of the natural language system. Due to structure and system differences existing between the levels developers have to use a special approach and a distinct formal language for each one of them. As a result, the problem of establishing transitions between levels becomes more complex which leads to less efficient solutions.

Formalized natural language knowledge is highly demanded in a whole range of applications, to name but a few: foreign language learning systems; Machine Translation; Semantic Search; Semantic Text Markup; Information and Knowledge retrieval from texts; Natural language interface for applied intelligent systems, etc.

We can assume that creating a unified base of linguistic knowledge (hereafter referred to as LKB, or the Linguistic Knowledge Base) will make the development of such systems a far simpler and faster process. Many developers even research teams are pursuing this goal, some are even fairly close to achieving it. This being said, there is still a large number of issues that haven't yet been solved or only partially solved. Following are several examples of such issues:

- Keeping the LKB up-to date, so that it reflects the current state of the language;
- Saving and adequate processing of phenomena that contradict the current norms (mistakes, dialects, and other individual features that exist in the production of every language speaker;
- Storing and efficiently processing large amounts of complex structured information about all language levels in the unified memory.

In this work we attempt to represent linguistic knowledge using a special homogenous semantic network based on a knowledge representation language called SC (Semantic Code, http://ostis.net). The creators of this language [Golenkov, 2001], [Ivashenko, 2009] believe that it will make possible to solve the above mentioned issues. The representations in the SC language will be hereafter referred to as the '**semantic model of the linguistic knowledge base**' (SMLKB). This model has some particularities including:

- It offers a unified representation of all language levels and stores corresponding information in the unified complex structured knowledge base;

- It represents linguistic knowledge using non-linear graph structures that are best suited for corresponding human cognitive models;
- It enables a formalized description of many conclusions based the discussion of the properties of objects or phenomena belonging to a given subject area (in this work we are using a knowledge-based approach and the subject area is “Natural Language”).

The above statements don’t mean that we assume we are obtaining some significant results. In this paper we only provide an outline and suggest yet another formal approach to the creation of a linguistic knowledge base. Representing the language as a homogenous semantic network we are also staging an experiment that should confirm or show the falsity of our hypothesis which says that such a method of visualization of language knowledge is more efficient and straightforward (“semantic”) than that of linear texts.

We should also mention that this topic is being investigated by many researchers across the world, and experimental models are being created, among others in projects that belong to the “semantic web” category. The most frequently mentioned achievements in this direction are improvements made in the information retrieval systems for Internet. However many issues, unfortunately, remain unsolved. For instance in may 2013 Russian web search giant Yandex announced its new project named “Ostrova” (‘Islands’) (http://beta.yandex.ru/promo), which makes use of a new search technology (http://www.seonews.ru/analytics/yandex-ostrova-tehnologiya-interaktivnogo-poiska) with a main goal to help users to achieve their goals significantly faster. It’s been more than a year now, however the project is still at the beta stage. The most popular internet search engine Google uses an additional markup system to tag its search results which saves time for the by helping him to figure out what a site is about before even proceeding to it. Since 2012 Google uses Knowledge Graph - semantic technology and knowledge base to improve the quality of its search engine with semantic-search information gathered from various sources (http://www.google.com/insidesearch/features/search/knowledge.html).

SC language developers have launched their website http://ims.ostis.net where all the information is presented as a knowledge base that can be navigated by using special search terms understood by the semantic network.

In our opinion, creating and researching various efficient and visual information representation methods in the Internet and in a whole range of specific applications is one of the most important tasks at this stage. This task is particularly important for intelligent tutoring systems for foreign languages.

It is also worth mentioning that the external representations (aimed at the user) of the information are tightly connected with its internal representation inside the computer system’s memory. This is one of the reasons why a lot of efforts are directed nowadays at researching new formats for knowledge storage and processing.

Basic Remarks

In order to keep things simple and logical we will narrow this discussion to the creation of a LKB for an intelligent tutoring system for foreign language teaching (ITS for FLT) [Yeliseyeva, 2012]. We will also use a metaphor when referring to this objective, i.e., when formalizing the knowledge about natural language what we really do is actually teach the foreign language to the computer. We will use this metaphor when talking about different aspects of the structure and contents of the educational LKB building its semantic model.

The results of our reasoning will be presented by means of a special form of homogenous semantic network which is described using a knowledge representation language called SC (SC-code), the base language of the open-source project OSTIS (Open Semantic Technology for Intelligent Systems) [http://www.ostis.net]. Its aim is to create a popular semantic technology for component-based design of intelligent systems with various purposes.

The structures of the language SC are called sc-texts or sc-structures. Sc-structures consist of sc-elements, sc-nodes and sc-arcs being the basic ones. The SC language can represent information using 2 notations: 1) graphical – the knowledge is represented using graphs in the SCg-code (Semantic Computer Graphic Code). It is one of possible flat (2D) visualizations of sc-structures; 2) textual (linear).

SC language semantics are based on the set-theoretical relation of belonging, where an oriented sc-arc from sc-element *X1* to sc-element *X2* means that *X2* belongs to the set *X1*. Sc-elements may have identifiers that we will write down using italic.

Hereafter we will discuss the formalization of Russian and/or Belarusian languages but we shall not claim that the approaches suggested in this work are complete or universal. However, we do hope that many of the ideas and formalisms offered hereafter, with necessary precisions and additions will be useful for creating LKB of many other languages.

Our reasoning is based on the experience we gained when creating online learner dictionaries for Russian and Belarusian hosted on http://rus.lang-study.com and http://by.lang-study.com respectively. These sites are the testing ground where Belarusian State University students enrolled in a bachelor, master and doctoral program upload their content. Since within these projects mainly learner's dictionaries of foreign languages are created, the contents of the dictionaries and thematic word groups are determined by the communicative objectives. To a certain extent these resources are being developed and extended spontaneously which, we have to admit reduces the quality of the results achieved. On the other hand, such projects tend to be constantly modified and improved as the competences of their creators improve. After all, their main objective is to train qualified professionals capable of creating linguistic resources rather than to produce a final product.

We would like to add that the present work is a new step in the development of these projects and should provide a foundation for creating ITS based on the aforementioned sites.

Definitions

First we shall clarify the meaning of the term 'linguistic knowledge base'. For that, a definition of "knowledge" by Gavrilova will be useful: "**Knowledge** is well-structured data" [Gavrilova, 2001].

Thus, we will define linguistic knowledge base as well structured data about the language.

Another useful concept is that of language structure [Kobozeva, 2009], which includes the dictionary (vocabulary) and the grammar. Thus, structured language data should include information about the vocabulary (lexis) and grammar of the language. Thus, we can subdivide the linguistic knowledge base in two components:

1) **lexical knowledge** (structured data); 2) **knowledge** (structured data) **about grammar**.

Hereafter the above elements of the LKB are considered in more detail.

As we've already mentioned above, the results of our reasoning will be presented in the form of sc-texts. In order to create such descriptions, the OSTIS technology requires the identification and comprehension of the notions and relations of the knowledge domain being formalized. This enables the creation of a sc-sublanguage that corresponds to the given knowledge domain and consists of sets of sc-elements (mainly, sc-nodes). The identifiers of such sc-elements are signs of concepts, relations and attributes identified in the knowledge domain. Therefore we will refer to the collection of resulting sc-texts as well as the description of the ontology of identified concepts and relations as **the semantic model of the linguistic knowledge base** (SMLKB).

Lexical Knowledge

It is well known that in order to master the vocabulary of a foreign language one needs not just the words with their translations in his native language, but also additional information about these words. Such additional information includes: 1) lexical meaning (meaning explanation) of the word; 2) grammatical properties and inflection rules; 3) information about word compatibility; 4) semantic and other types of relations; 5) particularities and examples of use: word combinations, sentences, texts.

Note that grammatical knowledge is also included in this list (cf. items 2 and 3). Thus, we don't separate the lexis from grammar. Moreover, in the present work we intentionally avoid using the common approach to the identification of language levels (morphology, syntax, semantics, etc.). As we've already mentioned, we are generalizing the experience of the creation of Russian and Belarusian dictionaries for http://rus.lang-study.com and http://by.lang-study.com. Besides, we also use the approach to creation of explanatory combinatorial dictionary described in the works of Melchuk [Melchuk, 1974]. In both cases we have noticed that solving lexicographic issues leads to results that can be applied to many other areas besides the creation of dictionaries.

Let's elaborate on the remarks we've made concerning the vocabulary of natural language (NL) as well as the possible ways of its formalization using the SC language or the experimental hypertext model of the aforementioned online learner dictionaries.

Lexical Meaning (meaning explanation) of a Word

Works on lexical semantics [Apresyan, 1995], [Kobozeva, 2009] suggest various ways of describing the meaning of a word. One of the approaches involves forming semantic fields. Componential analysis is carried out to define the hierarchy of topic groups as well as to match each word of the language with one or several of these groups. Many dictionaries, including semantic [Shvedova, 1998, 2002], ideographic and thematic dictionaries and thesauri are created in this way. An example of Russian language classification has been suggested in the semantic dictionary under the general editorship of Shvedova which can be accessed online at http://www.slovari.ru/default.aspx?s=0&p=2672. Learner's dictionaries at http://rus.lang-study.com and http://by.lang-study.com can be systematized by determining thematic word groups.

It's worth mentioning that at present there is no universal approach to the problem of defining the contents and the limits of semantic fields. In our opinion, that doesn't even seem to be possible, taking into account the objective nature of the natural language and the subjectivity of human perception. One thing is certain though: structures that are being used to systematize vocabularies of various languages are complex and non-linear. This is why we chose the SC language of semantic networks designed for describing and processing this type of non-linear structures.

Now let's consider a simple example. Fig.1 shows a fragment of a sc.g-text describing a series of words related to "human" – which is one of the topics («Человек») presented on the website http://rus.lang-study.com. Fig.1 A) displays a simplified structure where many nuances of Fig.1 B) are simply omitted. Let's take a closer look at both fragments of the semantic network. Arcs leaving the *части тела* (=parts of the body) sc-node fully reflect the basic semantics of the part-of relation described above. Indeed, the head (sc-node with the *голова* identifier), hand (*рука*), leg (*нога*) – all these are parts of the body, i.e., they belong to a set marked with the *части тела* sc-node. Semantics of all the sc-arcs leaving the sc-node *внешность* (=appearance) is similar. Naturally, the arcs leaving the *человек* (=human) sc-node should have different semantics and define other type of relation. However, it is not reflected in the Fig. 1 A). The arcs present in this figure can be understood so that the *parts of the body*, *appearance* and other sc-nodes belong to the *human* set. This description is obviously not really correct. That's why in the figure 1 B) we use a different style for the sc-arcs leaving the *человек* sc-node. These are so-called oriented couples which depict an oriented binary relation between the given sets. The semantics of the given relation are described by the arcs leaving the sc-node with the *включение множеств** (=inclusion of sets) identifier. This way we show that the *части тела*, *внешность* and some other sets are subsets (and not elements) of the *человек* set.

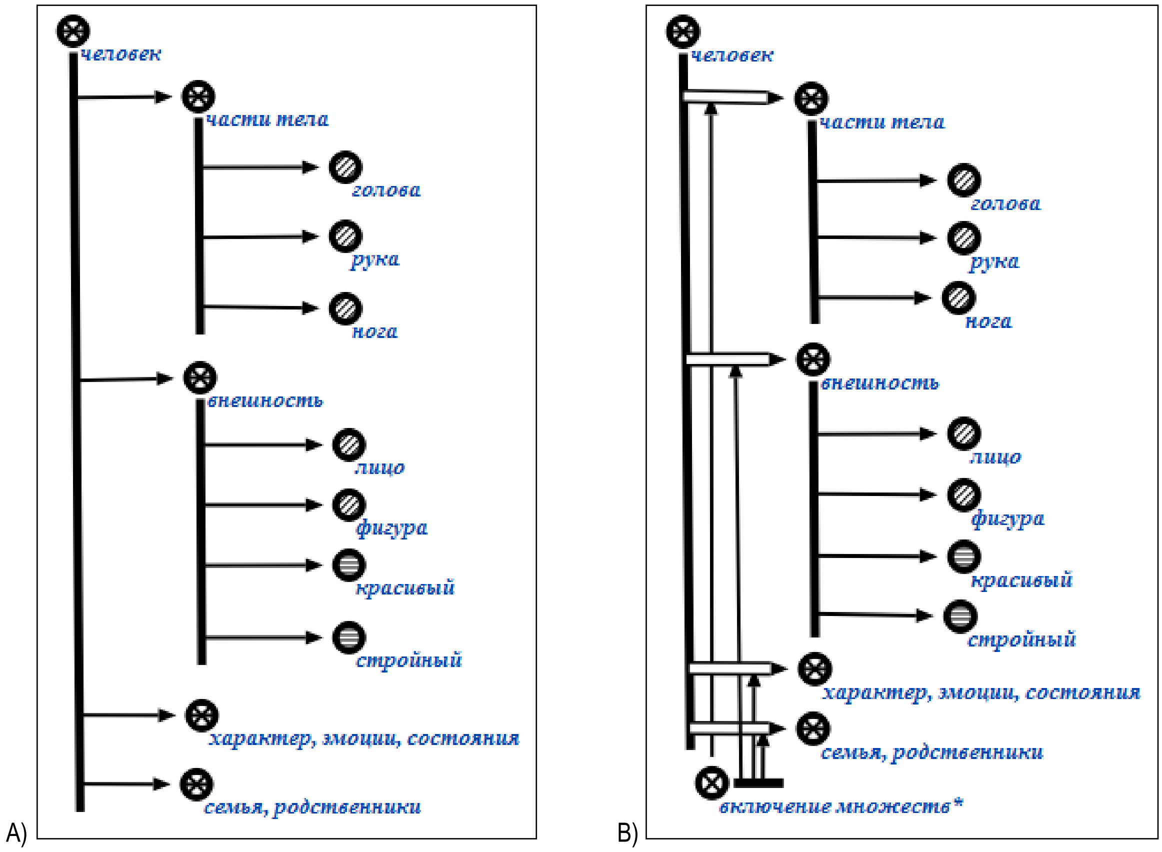

Figure 1. An example of thematic groups description in the dictionary

Fig.1 also uses various images of sc-nodes. These are extended possibilities of the SC language for describing the semantics of corresponding elements. For instance, the sc-node defines sets (groups) of elements of the same nature. In Fig.1 we have words that have in their meaning one common semantic marker. Using sc-nodes of type and subject (constant) elements of a subject domain (concrete, tangible and abstract, intangible respectively). No sc-arcs usually leave such sc-nodes. Finally, the sc-node in Fig 1 B) defines a relationship. Besides the relations we can use special sc-nodes to show attributes ( ) and a bunch of relations ( ). SC language graphical notation offers one more possibility to place elements in a more a more straightforward way. It's the so-called sc-tire – a thick line leaving the sc-node that becomes longer. This allows for additional arcs to be drawn to and from this sc-node without overloading the picture and making already complex networks even more confusing.

Let's consider the LKB fragment further. In order to make a more complete description of semantics of all sc-nodes shown in the Fig.1 we need to introduce two more set signs – *тематическая группа* (=thematic group) and *лексема* (=lexeme). Let's describe the corresponding statements about element typology using a linear sc-text notation:

тематическая группа -> *части тела*; *внешность*; *характер*, *эмоции*, *состояния*; *семья*; *родственники*;;

лексема -> *часть тела*; *голова*; *рука*; *нога*; *лицо*; *фигура*; *красивый*; *стройный*;;

Obviously some of the lexemes can be included in several thematic groups simultaneously. For instance, the word *лицо* (=face) can belong to *части тела* and to the *внешность* (Fig. 2).

Besides the attribute *характеристика_* (=property) in Fig. 2 indicates an additional semantic marker of the words *красивый* (=handsome) and *стройный* (=slender).

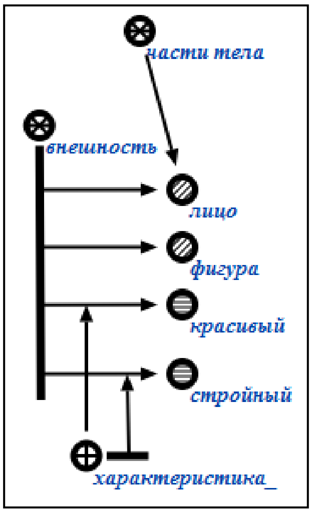

Figure 2. Additional descriptions the dictionary fragment

We will mention once more that we are not claiming that our reasoning is linguistically complete and precise. Our task is to consider in detail formalisms used for describing the semantic model of linguistic knowledgebase. We are well aware that many of our statements may be criticized by linguists and we fully accept this fact.

Of course the above mentioned method of describing word semantics of a natural language by assigning words to various semantic fields (thematic groups) is not a silver bullet. Componential analysis of lexical meaning consists in defining several semantic markers that together form this meaning. Linguists call these markers semes and distinguish at least three types of semes: differential seme, archiseme, and contextual seme. Fig. 3 shows an sc-text describing the lexeme *father* using the relation of *семантическая декомпозиция** (=semantic decomposition). For this description we've used the article «Сема» (Seme) from the linguistic encyclopedia accessible online at (http://tapemark.narod.ru/les/437c.html).

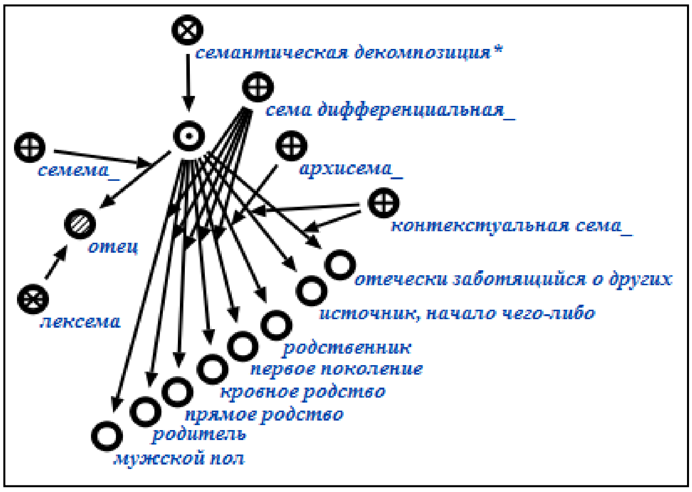

Figure 3. Description of lexeme father through enumeration of various semes

Let's mention here yet another particularity of the SC semantic network language we're using. The sc-nodes used in descriptions having same identifiers undergo the operation of joining when they are loaded into the sc-memory (also called graphodynamic by developers). Thus if we are an sc-node with the same identifier in different fragments of an sc-text, we are basically expanding the description of the same element of the semantic network. When creating a knowledgebase we mostly identify sc-nodes, so every time we mention the same node in our descriptions we basically add sc-arcs leading to or leaving this node.

In accordance with the above mentioned possibilities of the SC language, when similar semantic descriptions of the lexeme *mother* are added to the knowledgebase, sc-nodes like *parent*, *direct relationship*, *real kinship* and others will be completed with new sc-arcs leading to and from them. Further if we make a more complete description of a number of natural language words we will obtain quite a complex structure clearly showing semantic connections between these words. Using the connections between the sc-nodes defining semes finding semantically close words are relatively easy to find by certain markers (semes). Adding detail to the semantics of words that have already been described in such a way is as easy as adding missing connections (arcs) to the sc-node that represents the target lexeme.

Using the above mentioned method one can describe almost any kind of information involving semantics and functions of words in the language. To do that we need to introduce and describe respective relations and their attributes. We will add several examples in what follows.

Grammatical Properties and Inflection Rules

In Fig. 4 we see a fragment of an sc-text describing grammatical properties of several words that have been described above in the context of a hyerarchical structure representing semantic field in the Fig.1. Unlike in Fig. 1, we've used a slightly different depiction of sc-nodes, more specifically, rectangular boxes with words inside. These are the so-called text (string) contents of sc-nodes. To avoid unnecessary complexity we don't display the identifiers of these nodes since in this case they are identical with the contents. Thus, in a general case identifiers are not necessary and may only be needed to ensure more precise merging of sc-nodes when they are being loaded into the graphodynamic memory. We will skip the textual explanation of the sc-structure in Fig. 4, assuming that is quite straightforward.

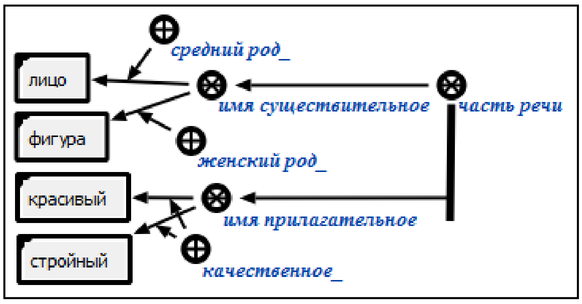

Figure 4. Description of grammatical properties

Up until now we've been describing different LKB at the declarative level. So, for instance, in the Fig. 5 we've described the inflectional paradigm of the word *аудитория* (=lecture hall) using an sc-text at a declarative level (cf. the dictionary article at http://rus.lang-study.com/slovar/100-slov-dlya-1-kursa/auditoriya-3).

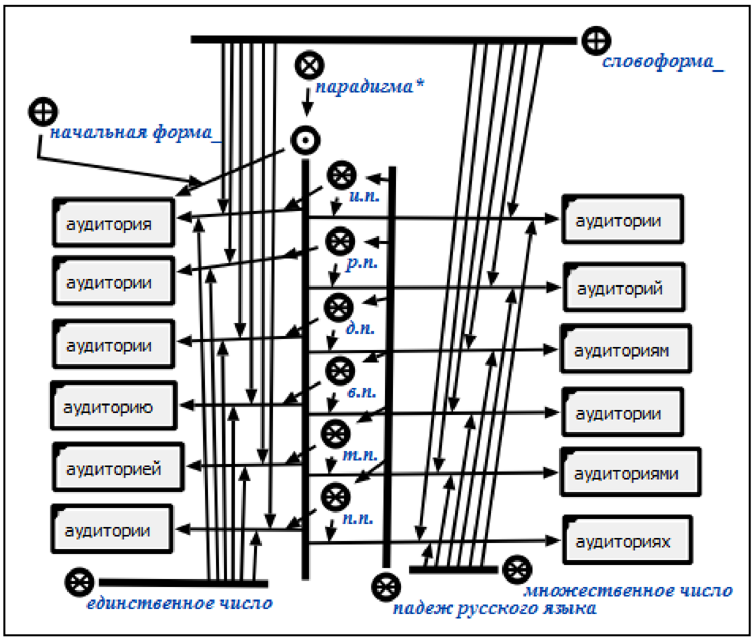

Figure 5. Description of the word inflection paradigm

At the procedural level of knowledge the knowledgebase should describe the corresponding rules of inflection, rules for automatic identification of grammatical properties of the input word forms, etc. In this work we are not attempting to provide such descriptions because it would require a lengthy explanation of the functionality of the SC language, of the processing operations used by the graphodynamic memory [Golenkov, 2001], as well as of the procedures of automatic analysis and synthesis. Taking into account the current state of the research we don't consider such exhaustiveness necessary, since the number of available works on automatic text processing nowadays is impressive. The tools of the OSTIS project used in this research allow for combination of different types of knowledge including external procedures and algorithms in one system.

Here we will only mention some examples of use of the suggested LKB descriptions for problem solving. The Fig. 6 shows an example of an sc-text describing the search for a word form using a given lexeme (*аудитория*) and a set of grammatical properties (*singular*, *Genitive Case*). This description makes use of sc-variables defined with squares instead of circles for sc-nodes or with dashed lines for sc-arcs. Pattern-matching search is a basic operation of information retrieval used by the sc-machine, initialized using respective sc-descriptions and semantic network search procedures – all stored in the sc-memory.

Fig.7 shows an sc-text describing the search of values for the grammatical properties of *case* and *number* of the given word form *аудитории* and its initial form.

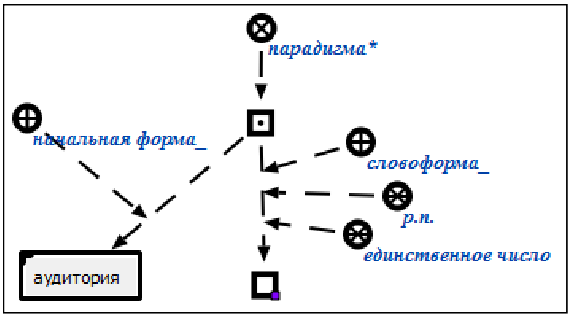

Figure 6. Example of word form search based on the inflectional paradigm

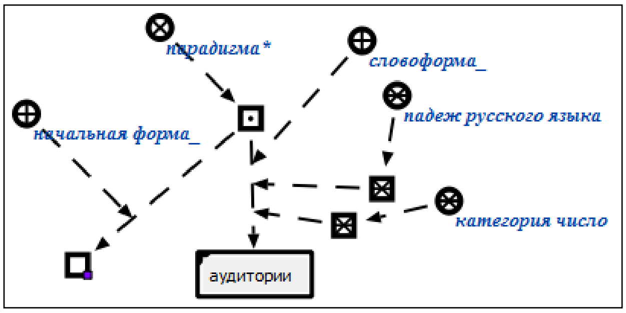

Figure 7 Example of search for grammatical properties of a word form

Information About Word Compatibility

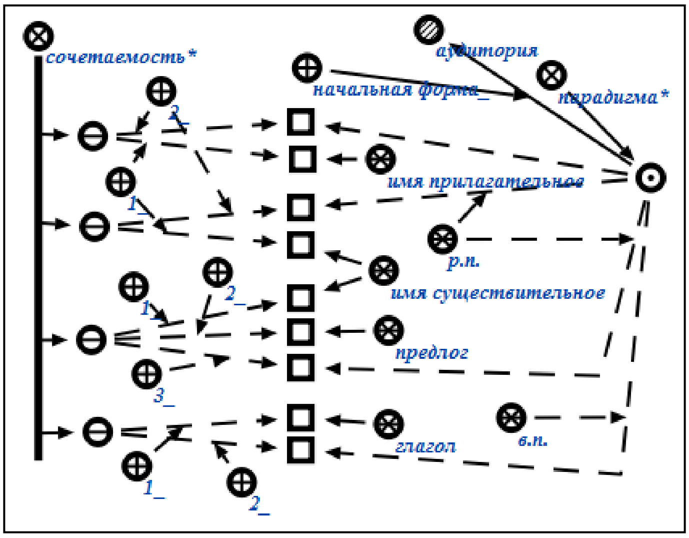

Figure 8. Description of information about the compatibility of a lexeme with other words

To help students develop the ability to use words in combinations and complete sentences and, as a result, to ensure effective communication many foreign language textbooks offer information about the compatibility of new words with other lexical units. For instance, when teaching Russian as a foreign language one can use specific formulas, some of which can be found in the dictionary article «Аудитория» at rus.lang-study.com (http://rus.lang-study.com/slovar/100-slov-dlya-1-kursa/auditoriya-3). The Fig. 8 shows a fragment of description of some of these formulas. For this we have introduced the *сочетаемость** (=compatibility) relation, oriented bundles of which form templates of a certain type (the sc-construction contains variables) which, when supplied with actual word forms that have specified grammatical properties, will provide appropriate word combinations. In this example, containing many sc-variables we have intended to demonstrate the most general form of description and the morpho-syntactic compatibility. If some variables are replaced with actual sc-nodes (lexemes) and/or word forms we will obtain an example of the description of lexical compatibility in the LKB [Popov, 2004]. Finally if similar structures also indicate semes and the thematic groups (semantic fields) or other semantic markers of lexemes, then we will obtain a rough description of semantic compatibility.

The use of such descriptions in the functioning of an application system seems sufficiently obvious. Appropriate fragments of the semantic network along with sc-variables will serve as a model for searching and/or generating necessary information in the sc-memory.

The above suggested description is not the only and perhaps not the most optimal method. Many aspects of lexical compatibility, in particular, can be described using the lexical functions, suggested by Melchuk in the framework of the Meaning ⇔ Text model [Melchuk, 1974]. Some examples of such descriptions will be provided further.

Semantic and Other Types of Relations

Let's consider the possibilities that our SMLKB offers for describing various relations between lexemes.

Fig.9 and Fig.10 present examples of description of the synonymic relation. To indicate such a relation in the SC language we use the abbreviation *Syn**, suggested by Melchuk in his Meaning ⇔ Text model [Melchuk, 1974]. In Fig. 10 the type of synonymic relation between lexemes is specified using the attribute *экспрессивно-стилистическая_* (=expressive-stylistic). Besides, in Fig.10, a symmetrical bundle is defined by a double line to increase the visual clarity. This is supported by the graphical notation of the SC language.

The synonymic relation is used more widely in the SC language. In particular, it is used to describe interlingual synonymy (Fig. 11). In some cases, however (for instance, when there is no exact equivalent), it is more appropriate to use the sc-relation of *трансляция** (=translation) (Fig.12).

Figures 13 and 14 show graphic depiction the relation of hyponymy (based on the example by [Kobozeva, 2009]).

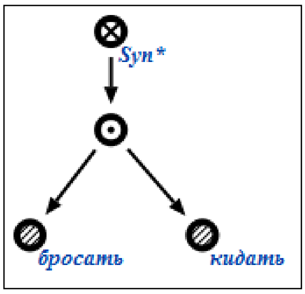

Figure 9. Describing a synonymic relation

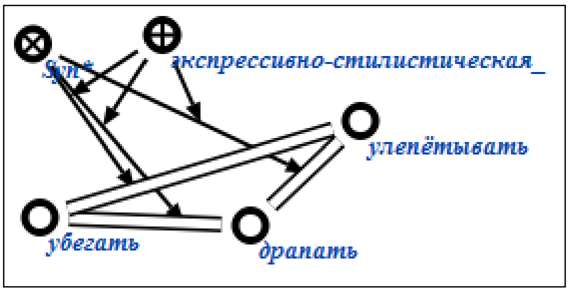

Figure 10. The description of an expressive type of synonymic relation

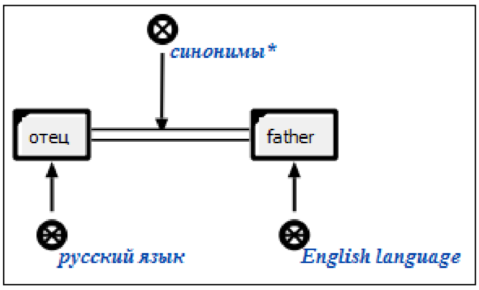

Figure 11.The description of the interlingual synonymy

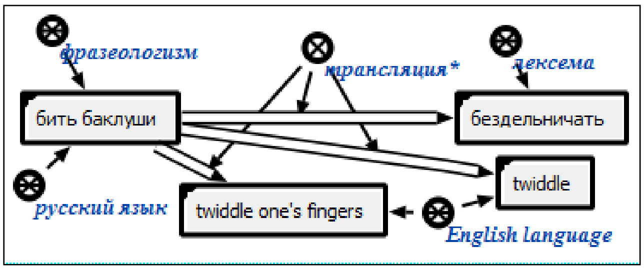

Figure 12. Using the translation relation

In the same manner we could describe various kinds of relations and lexical functions between separate words, word combinations and phrases. No less important is, for instance, the description of associative (http://wordassociations.ru/) and some other relations.

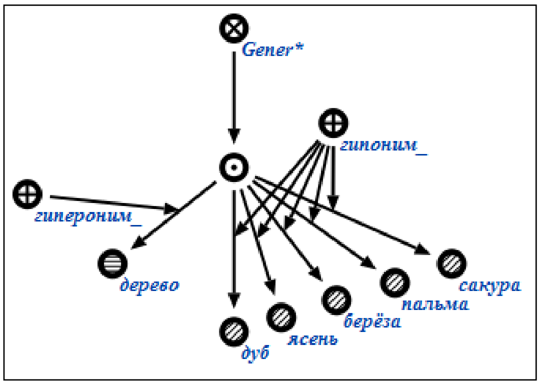

Figure 13. Describing the relation of hyponymy

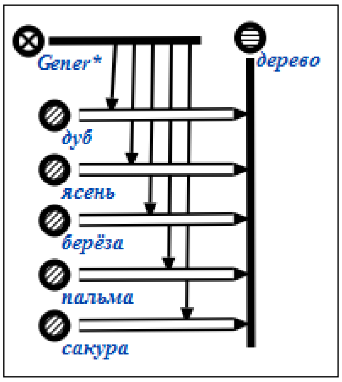

Figure 14. An alternative visualization of the hyponymy

The Figure 15 shows an example of relation “whole – part” tagged using the identifier *Sing** , borrowed from the [Melchuk, 1974]. The contents of this example have been borrowed from [Kobozeva, 2009].

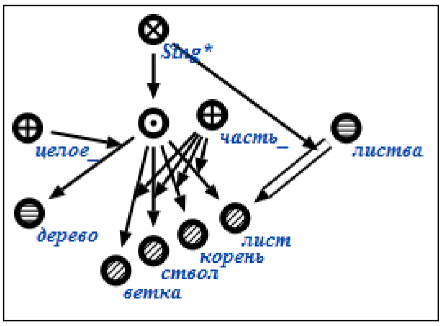

Figure 15. The description of the “whole – part” relation

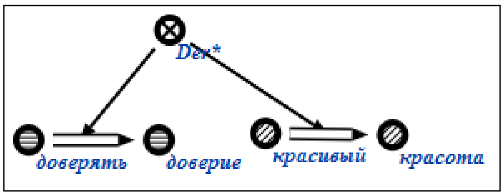

Figure 16. Examples of derivational relations

Finally, in the Fig.16 we provide an example of a derivational relation (*Der**). We can use additional attribution to specify particular derivation methods or appropriate rules to enable the function of automatic generation, but won’t do that due to space limitations.

We have started systemizing some of the relations described above in the dictionary article sat http://by.lang-study.com by determining appropriate fragments with subheadings (cf. http://by.lang-study.com/slounik1/ezha-harchavanne/sadavina1/yablyk).

Particularities and Examples of Word Use in Combinations, Sentences and Texts

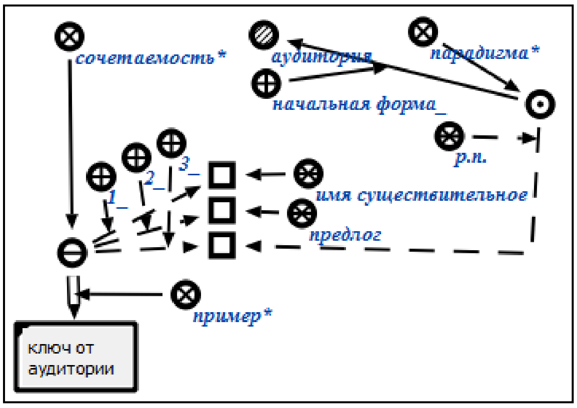

Figure 17. Describing an example of use along with as the compatibilityFig 3. Description of lexeme father through enumeration of various semes

Some particularities of word use in speech have been mentioned by us before during the discussion on the compatibility and other relations and lexical functions. Here we will provide example of fragment of sc-text that contain explicit references to instances of word use (Fig. 17).

Such descriptions are especially useful for an instructional LKB of an ITS. For instance, on our websites such examples represent the essential information on a lexeme. Incorporating such examples along with their most adequate translations in the LKB of machine translation systems can be used to improve the quality of translation.

Knowledge (structured data) About Grammar

As we've mentioned before, while formalizing the vocabulary of a language we also described some aspects of its grammar. Such aspects include grammatical properties (Fig.4), inflectional paradigm (Fig.5), information on compatibility (Fig.8) etc. In this work we discuss the formalization of knowledge of Russian language which has a complex morphology. Professionals that deal with the lexical semantics note that Russian doesn't have a clear distinction between the lexis and the grammar [Kobozeva, 2009]. Therefore one may get an impression that the LKB described in this work pays too much attention to the morphological level. At the syntactical level, the formal description of natural language texts through sc-texts is done in the same way. So, for example, in [Yeliseyeva, 2014] shows a somewhat simplified description of the sentence structure at the surface syntax level. Moreover is shown there the correspondence between the text in a natural language and its translation in the SC language.

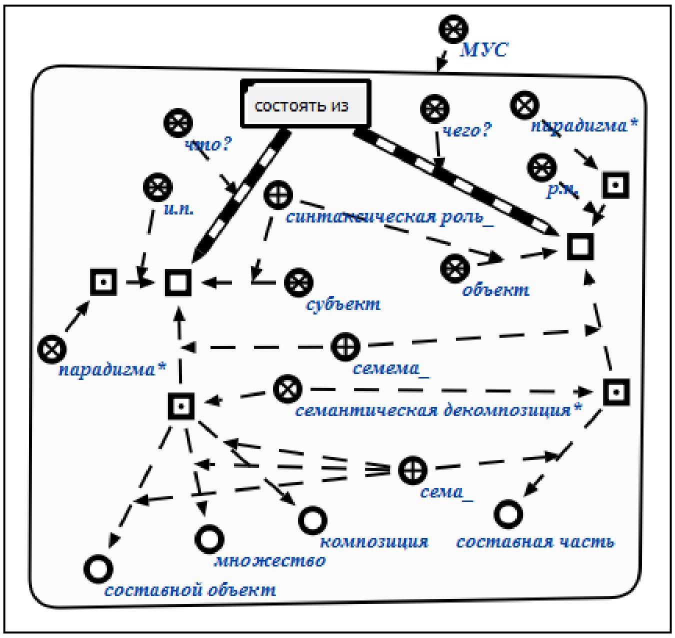

Figure 18. A description of a model of coordination of a word

To execute automatic synthesis and analysis of sentences in a natural language, sc-structure templates containing in their descriptions sc-variables should be stored in the LKB or generated in the sc-memory. Models of word coordination (MWC, [Apresyan, 1995]) seem to be good source data for creating such templates. In Fig. 18 we provide an example of a formal description of the MWC of the lexeme *состоять из* (=consist of). At the same time this MWC provides morpho-syntactic information as well as the semantics of possible lexemes that are coordinated by the lexeme *состоять из*. For the sake of simplicity we use coordination questions *что?* и *чего?* instead of specifying semantic valence of the lexeme *состоять из*.

In the functioning of a natural language application system formal representations of MWC can be used for analyzing input sentences as well as for generating texts in a natural language.

Conclusion

In the present work we have attempted to describe the semantic model of a linguistic knowledge base in the format of homogenous semantic networks organized in a particular way using the SC language of knowledge representation. One of the particularities of the suggested model is that it provides a uniform presentation of all language levels and stores the corresponding information in the unified knowledgebase. This has become possible because we intentionally didn't define separate levels in the process. In our opinion, this approach allows us to unify the mechanisms of automatic processing of natural language structures. Based on the present model it becomes possible to create applied natural language systems organized in accordance with the principles of knowledge-based systems. Besides, in future natural language knowledge can become the foundation for knowledge bases in other domains of knowledge [Yeliseyeva, 2011]. We also consider that the LKB will become an important tool for deep research of the natural language.

Due to the size limitations of the article we cannot provide here all the relations but we hope that we've managed to present the most general approaches to such descriptions.

For applied intelligent systems, LKB is the basis for an efficient realization of natural language interfaces. The suggested approach also enables the realization of intelligent tutoring systems for Russian language teaching. The present work uses the concept of semantic analysis of the Russian language structure with an outlook for the best practices in its teaching.

Bibliography

[Shannon, 1949] C.E.Shannon. The Mathematical theory of communication. In: The Mathematical Theory of Communication. Ed. C.E.Shannon and W.Weaver. University of Illinois Press, Urbana, 1949.

[Apresyan, 1995] J.D. Apresyan. Selected Papers, Vol. 1. Lexical Semantics: 2nd ed., corr. and add. – M.: School "Languages of Russian Culture" publishing company "Eastern Literature" RAS, 1995. (rus)

[Gavrilova, 2001] Gavrilova T.A., Horoshevsky V.F. The knowledge bases of intelligent systems. St.-Petersbourg, 2001. (rus)

[Golenkov, 2001] Golenkov V.V., Yeliseyeva O.Y., Ivashenko V.P. et al. Representation and processing of knowledge graf-dynamics associative machines: Monograph; BSUIR, 2001. (rus)

[Ivashenko, 2009] Ivashenko V.P. Semantic technology of engineering knowledge bases // BSUIR's Papers. – 2009. - No.7. – P.44-51. (rus)

[Kobozeva, 2009] Kobozeva I.M. Linguistic semantics: guide. 4th ed. – M.: Book House "LIBROKOM", 2009. (rus)

[Melchuk, 1974] Melchuk I.A. Experience of the theory of linguistic models «Meaning ⇔ Text». – M.: Science, 1974. (rus)

[Popov, 2004] Popov, E. V. Natural language interaction with the PC / E. V. Popov. – Moscow. : Yeditorial URSS, 2004. (rus)

[Yeliseyeva, 2011] Yeliseyeva O.Y. The use of artificial intelligence technologies in language teaching // Open Semantic Technologies for Intelligent Systems (OSTIS-2011) – Minsk, BSUIR, 2011. – P. 363 – 370. (rus)

[Yeliseyeva, 2012] Yeliseyeva O., Kim Y. Intelligent Tutoring System for Belarusian as a Foreign Language // Artificial Intelligence Driven Solutions to Business and Engineering Problems : Galina Setlak, Mikhail Alexandrov, Krassimir Markov (Eds.) - I T H E A®, Rzeszow, Poland – Sofia, 2012, Bulgaria. – PP. 93 – 101.

[Yeliseyeva, 2014] Yeliseyeva O.Y. Semantic conceptual design of natural language interface of intelligent system // Open Semantic Technologies for Intelligent Systems (OSTIS-2014) : V.V.Golenkov at al (ed.) – Minsk: BSUIR, 2014.

[Shvedova, 1998, 2002] Russian semantic dictionary. Dictionary, systematized on the word classes and values: In 6 vol. / N.Y. Shvedova (ed.). - RAS. Russian Language Institute, M.: Azbukovnik, 1998, 2002. (rus)

[http://www.ostis.net] URL: http://www.ostis.net – an open source project Open Semantic Technology for Intelligent Systems.

Authors' Information

Olga Yeliseyeva – *PhD, associate professor of the Department of Applied Linguistics, Faculty of Philology at the Belarusian State University, 51 Napoleona Ordy street, appt.118 Minsk, P.O. Box: 220045, Belarus, e-mail: volga.eliseeva@gmail.com*

Major Fields of Scientific Research: E-learning, Artificial Intelligence, Natural Language Processing

Kim Yury – *Master course student at Yonsei University, Seoul, 28 Tikotskogo street, appt. 83 Minsk, P.O. Box 220119, Belarus e-mail: 6rikim@gmail.com*

Major Fields of Scientific Research: Corpus Linguistics, Information Retrieval.