SK-languages as a Powerful and Flexible Semantic Formalism for the Systems of Cross-Lingual Intelligent Information Access

Vladimir A. Fomichov
School of Business Informatics, Faculty of Business and Management
National Research University Higher School of Economics (HSE), Kirpichnaya str. 33, 105187 Moscow, Russia
E-mail: vfomichov@hse.ru, http://www.hse.ru/eng/org/persons/67739

Keywords: semantic parsing, theory of K-representations, abstract meaning representation, formal representation of semantic content

Received: January 26, 2017

The first starting point of this paper is the broadly accepted idea of employing, as a promising methodology, an artificial semantic language-intermediary for the realization of automatic cross-lingual intelligent information access to natural language (NL) texts on the Web. The second one is the emergence in computational semantics during 2013-2016 of great interest in the semantic formalism (more exactly, notation) called Abstract Meaning Representation (AMR). This formalism was introduced in 2013 in an ACL publication by a group consisting of ten researchers from UK and USA. This paper shows that much broader prospects for creating semantic languages-intermediaries in comparison with AMR are opened by the theory of K-representations (TKR), developed by V. A. Fomichov. The basic mathematical model of TKR describes the regularities of NL structured meanings. The mathematical essence is that this model introduces a system consisting of ten partial operations on conceptual structures. Initial version of this model was published in 1996 in Informatica (Slovenia). The second version of the model (stated in a monograph released by Springer in 2010) defines a class of formal languages called SK-languages (standard knowledge languages). It is demonstrated that SK-languages allow us to simulate all expressive mechanisms of AMR. The advantages in comparison with AMR are, in particular, the possibilities to construct semantic representations of compound infinitive constructions (expressing goals, commitments, etc), of compound descriptions of notions and sets, and of complex discourses and knowledge pieces.

Povzetek: Opisani so SK-jeziki za fleksibilno med-jezikovno dostopanje.

1 Introduction

During last decade, one has been able to observe a quickly growing interest in the design of computer intelligent agents fulfilling cross-lingual information retrieval (CLIR) on the Web. It is a consequence of emerging a huge, permanently increasing number of Web-sources in languages being different from English. In September 2012, a seminar on Multilingual Semantic Web (MSW) was organized in Germany in the Dagstuhl Castle. The proceedings of this seminar contain the following data [5]: in the year 2010, the number of non-English-speaking Internet users was three times higher than the number of English-speaking users (1430 million vs. 536 million users). That is why the problem of developing an MSW is very topical [24-26, 35, 49, 56].

It is broadly accepted that a promising approach to the realization of CLIR on the Web is employing a special semantic language-intermediary (SLI) in order to represent in the same format both semantic content of a user query and semantic content of the analysed fragment of a text in natural language (NL) [4, 7, 13-20, 24-26, 30, 32, 46, 49, 51, 52, 56].

The problem of creating a broadly applicable and flexible SLI goes far beyond the scope of CLIR. During last decade, the semantic parsing branch of computational linguistics has been considerably strengthened and expanded [36]. The main objective of this branch is to develop and implement the algorithms extracting meanings from NL-texts and forwarding them to pragmatic subsystems of applied intelligent systems. The real resurrection of semantic parsing branch (after two decades when statistics-oriented approaches to NL processing dominated) has been caused, first of all, by the stormy progress in designing autonomous intelligent agents (robots) and various mobile devices (cell telephones, planchettes, etc.) [36, 44, 45]. Another reason is the problem of understanding Web-sources in many natural languages on requests of the end users or on requests of computer intelligent agents. The use of SLI is also reasonable in full text question-answering systems and in NL interfaces (in particular, to robots and mobile devices) even in case of the texts in one language.

There is one more circumstance showing high topicality of developing broadly applicable and flexible SLIs. During last decade, several IT-companies have emerged in different countries whose principal objective is to combine the informational technologies of Semantic

Web and NL processing. In particular, these are Ontos GmbH in Swizerland [40, 53] and Cambridge Semantics Inc., The Smart Data Company in Boston, MA, USA [6].

During last decade, many scholars have seen a reasonable way of creating preconditions of understanding NL-texts by computer systems in developing special linguistic databases containing sentences associated with manually constructed semantic representations (SRs); in other terms, associated with semantic annotations. Since the year 2013, numerous papers have been published devoted to employing the notation called Abstract Meaning Representation (AMR) for constructing semantic annotations of NL sentences, in particular, of sentences in English, Czech, and Chinese [1, 2, 35, 43, 47, 54, 55].

The aim of this paper is to attract the attention of the researchers in computational semantics to the fact that there is a formal theory opening *much broader prospects* for building SRs of NL sentences and discourses in comparison with AMR. It is the theory of K-representations (knowledge representations) - an original theory of designing semantic parsers of NL-texts with the broad use of formal means for representing input, intermediary, and output data of the algorithms. Besides, it enriches the logical-informational foundations of MSW, multi-agent systems, E-commerce, knowledge representation in advanced ontologies, and knowledge representation in multi-media databases. The monographs [21, 25] state two versions of the theory of K-representations (TKR). It is an expansion of the theory of K-calculuses and K-languages (the KCL-theory). The basic ideas and results of TKR are set forth in numerous publications both in Russian and English, in particular, in [12-30]. TKR is the kernel of Integral Formal Semantics of NL, its basic principles and composition are stated in [16] and in Chapter 2 of [25].

The structure of this paper is as follows. Section 2 analyses related approaches, the main attention is paid to Semantic Role Labeling, Frame-Semantic Parsing, and Abstract Meaning Representation. Section 3 contains a task statement. Section 4 shortly describes the expressive mechanisms of SK-languages, introduced by TKR. Section 5 sets forth principal distinguished features of the algorithms of semantic parsing proposed by TKR. Section 6 shortly indicates the computer applications of TKR. Section 7 outlines the prospects of using SK-languages in the development of MSW. Section 8 concludes the paper.

2 Related approaches

2.1 Semantic role labeling branch of computational semantics

The goal of extracting meaning from NL-texts (and constructing its complete or partial representation) emerged in many application domains in early 2000s and initiated a number of research projects throughout the world. The main stream in this field includes, in particular, the interrelated branches called Semantic Role Labeling (SRL) and Frame-Semantic Parsing (FSP). The principal task considered in SRL is to find semantic relations (called *semantic roles*) between the verbal forms (and some other predicate words) and the dependent word groups. For instance, it is possible to find semantic roles *Agent*, *Phenomenon,* and *Time* in the sentence "The Russian Nobel laureate Ivan Pavlov discovered conditional reflexes in the beginning of the XXth century".

The aim of SRL is, firstly, to find realized semantic roles and, secondly, to construct a formal expression called semantic representation in order to process it in the context of a discussed situation and an ontology. The fundamental problem of SRL is that in early 2000s one felt the lack of formal means allowing for reflecting semantic structure of arbitrary sentences.

Example. Let S1 = "Yesterday Robert heard that the firm "Rainbow" would move to Manchester", S2 = "Robert decided to leave the firm "Rainbow".

Regretfully, the field SRL as far as five years ago didn't possess effective formal means for building SRs of sentences with complex direct or indirect speech, with infinitive constructions, and with modalities. In particular, it applies to the sentences S1 and S2.

A significant binary event in the development of this branch was the publication of the pioneer work [31] on a computer program for statistical SRL and the creation of PropBank annotations depositary [33]. These two publications became the starting point for designing a number of applied computer systems aimed at finding predicate-argument structures reflecting semantics of sentences and short discourses.

The PropBank annotations consist of phrase-structure syntax trees from the Wall Street Journal section of Penn Treebank [38] complemented by predicate-argument structures for the verbs. The PropBank uses core roles *ARG0* through *ARG5*, and these roles have different interpretations for different predicates. There are many studies aimed at SRL and using PropBank conventions [3, 39, 42]. The problem with using predicate-argument structures is that the roles ARG2 – ARG5 serve many different purposes for different verbs [58].

A way out is provided by the branch of NL processing (NLP) called Frame-Semantic Parsing and closely connected with the branch SRL [9]. The basis of this branch is the linguistic resource FrameNet [10], it stores a significant information about lexical semantics and predicate-argument semantics of sentences in English. The FrameNet lexicon contains semantic frames, each of them includes a list of lexical units – associated words and word combinations that are able to evoke a considered semantic frame in an NL expression. Besides, each semantic frame from FrameNet indicates several roles corresponding to the facets of the scenario represented by the frame. One says that targets are the predicates (verbs, etc.) evoking frames and arguments are a word or a phrase filling a role.

For example, the frame JUDGMENT from the FrameNet database contains the hand-annotated sentence "She blames the Government for failing to do enough to help". In this sentence, the following semantic roles are distinguished: *Judge* in the pair (She, blames), *Evaluee* in the pair (blames, the Government), *Reason* in the pair (blames, for failing to do enough to help). In the FrameNet database, the considered sentence is represented as follows:

[*Judge* She] **blames** [*Evaluee* the Government] [*Reason* for failing to do enough to help].

In comparison with PropBank, containing verbal predicates, FrameNet includes not only them but also adjectives, adverbs, and prepositions.

2.2 Abstract meaning representation formalism

In late 2000s and early 2010s, it was possible to observe a serious incompleteness of the field SRL. As it was mentioned above, the principal objective of the studies on SRL was to develop the methods and algorithms aimed at discovering semantic roles realized in the sentences. The purpose of discovering semantic roles is the use of this information in building SRs of sentences and discourses for interacting with pragmatic subsystems of applied intelligent systems.

However, the scholars in the field SRL possessed only rather restricted formal tools for building SRs of sentences. First of all, they felt the lack of convenient formal means for building semantic images of compound objects' and situations' descriptions, of sentences with attribute clauses of purpose, of sentences with infinitive constructions, and of sentences expressing modalities. That is why in late 2000s and early 2010s the scholars looked for more expressive semantic formalisms.

As a result, a new attention has been attracted to the semantic formalism called Abstract Meaning Representation (AMR), it was introduced in [34]. This formalism began its new life in a modified form after the publication of the paper [1].

AMR of a sentence S is an acyclic, rooted, directed graph with special marks of the vertices and edges. According to [34], a mark of a vertex has the form (*label/concept*), where *label* is a mark of an entity (e.g., *label* = m1) and *concept* is a string of the form |wd1| or |wd1,…, wdk|, where wd1, …, wdk are the words or word combinations expressing one notion (examles: |dog|, |eat, take in|).

The paper [1] considers additional forms of concepts' descriptions: the framesets of the linguistic database PropBank ("want-01", etc.), special entity types ("world-region", etc.), the kinds of quantities ("distance-quantity", etc.), and logical connectives "and", "or".

It is possible to distinguish several main reasons for explaining the quickly increasing interest in AMR.

Reason 1. The possibility to explicitly indicate semantic roles in the descriptions of events. It should be noted that AMRs use generalized semantic roles *arg0*,…, *arg5* employed in PropBank framesets [1].

Example 1 [1]. The sentence "The man described the mission as a disaster" can be associated with the AMR

(*d/describe-01 :arg0* (*m/man*) *:arg1* (*m2/mission*) *:arg2* (*d/distance*)).

Reason 2. The possibility to build compound designations of various entities from application domains.

Example 2 [1]. The expression "a singing boy from the college" can be associated with the AMR

(*b/boy :arg0-of* (*s/sing-01*) *:source* (*c/college*)).

Reason 3. A way of describing semantic structure of sentences with infinitive constructions.

Example 3 [35]. Let T1 = "The boy wants to go to New York". Then T1 may have the following AMR:

(w/want-01 :arg0 b/boy :arg1 g/go-01
:arg0 b :arg1 c/city
:wiki "New York"
:name (n/name :op1 "New"
:op2 "York")).

Reason 4. The possibility to describe semantic structure of sentences with modal words and infinitives.

Example 4 [1]. The sentences "The boy doesn't have to go", "The boy isn't obligated to go", and "The boy need not go" may be associated with the AMR

(*p/obligate-01 :arg2* (*g/go-01*)
:arg0 (*b/boy*) *:polarity* -)).

Another reasons are the possibilities to describe semantic structure of (a) the questions with interrogative words; (b) noun groups (e.g., "Elsevier N.V., the Dutch publishing group"), (c) sentences expressing the conceptual qualification relation ("This woman is a lawyer", etc.).

It is possible to distinguish the following principal shortcomings of the AMR notation from the standpoint of using it in the models and algorithms of semantics-oriented NL processing.

1. Our linguistic intuition says that (a) the main words and word combinations of the sentences refer to various things, situations, and abstract entities; (b) there are various directed semantic connections between the fragments of the sentence, in particular, between such main words and word combinations. A directed graph with special marks of the vertices and edges is the structure visualizing quite well this perception of a sentence by our linguistic intuition. However, this product of scientific thought can be characterized as a surface, non deep penetration into the mechanisms of NL semantics. That is why the AMR notation makes only a rather small contribution to the creation of the models reflecting the essence of sentence understanding with respect to a knowledge base.
2. The linguistic intuition of the scholars (not only of linguists) having command of several natural languages (e.g., of Russian and English or of English, French, and German) says that there are several mental mechanisms underpinning the construction of NL semantic structures in different languages. For instance, English, Russian, French, and German do have infinitive constructions and compound descriptions of sets. However, the AMR approach doesn't formulate any conjecture about a system of expressive mechanisms being responsible for constructing mental representations of sentences even in one language – in English.

3. Due to the above said, the AMR approach doesn't give a special formal status to such constructions as semantic images of infinitive expressions, compound designations of sets, and of sentences with modality. That is why the AMR approach seems to be of small use for constructing semantics-oriented models of NL communication.
4. The group of general semantic relations used in AMR seems to be a huge bag containing, in particular, such relations of different kinds as *:age*, *:destination*, *:consist-of*, and *:purpose*. The first unit is the name of a function, the second – fourth units are the names of the relations being not functions. These principal peculiarities are not taken into account by the AMP approach.
5. The AMR approach says nothing about the SRs of discourses.

3 Task statement

It seems to be reasonable to analyse the new demands to computational semantics in the context of the problems faced by computational linguistics (CL) as a whole.

The analysis of many publications describing the projects on NLP shows the existence of a gap (very often, a huge gap) between the employed theoretical tools and the real demands of the studied problems. Let's consider only one example. The linguistic processor BLUE (= Boeing Language Understanding Engine) was developed as an advanced information processing tool for the Boeing company. The system is able to build SRs of sentences of many kinds. In first section of one of the papers describing BLUE the authors state that the system uses the formal means of first-order logic (FOL) for constructing SRs of sentences [8]. However, we get to know from the second section of the same paper that the system BLUE "allows propositions to themselves be arguments to other propositions as a nested structuring". For instance, the system constructs an SR of the sentence "The man wanted to leave the house".

This step immediately leads us beyond the scope of FOL. The reason is that atomic formulas of FOL can't include the arguments being formal semantic images of infinitive constructions ("to leave the house", etc.). That is why the Boeing system BLUE, in fact, has no adequate theoretical background.

Analysing the development of CL as a whole during last twenty five years, it is possible to observe a shift to numerous engineering projects for solving particular practical tasks and the lack of attention to fundamental studies.

It seems that one of the brightest descriptions of recent and current situation in CL is given by Dr. Shuly Wintner from Computer Science Department of the University of Haifa, Israel [57]. The starting point for Dr. Wintner was high appreciation of the role played by mathematical theories in the development of many branches of engineering. For instance, air dynamics underpins the design of airplanes, and hydrodynamics is the basis for constructing ships. In this connection the following questions were posed by Dr. Wintner:

"What branch of science underlines NL Engineering? What is the theoretical infrastructure on which we build our applications? And what kind of mathematics is necessary for reasoning about human languages?"

It would be very natural to expand this list of fundamental questions by means of adding the question posed in [36]: "How to formally represent the semantics of language?".

The need of developing a comprehensive formal framework for creating an MSW makes very up-to-date the question about mathematical foundations of computational semantics being the core of modern CL.

The analysis shows that the current state of computational semantics demands the development of an applications independent semantic formalism being convenient: (a) for describing semantic structure of sentences including, in particular, infinitive and gerundial (for English) constructions expressing the goals commitments, commands, wishes, etc, the attributive clauses of purpose, complex direct and indirect speech, compound denotations of notions and sets; (b) for presenting semantic structure of discourses, in particular, of discourse with the references to the meanings of previous sentences or larger fragments of the text; (c) for building representations of knowledge pieces, including the definitions of notions; (d) for constructing formal representations of simple and compound goals of people, robots, and organizations.

This combination of expressive mechanisms is not proposed by FOL, Discourse Representation Theory, Theory of Conceptual Graphs, Episodic Logic [48], and Abstract Meaning Representation.

It is also possible to look at the formulated task from a more general position. The analysis of the scientific literature on semantic parsing and an MSW provides serious arguments in favour of putting forward the following conjecture*: it is high time for creating a new paradigm* for considering numerous theoretical problems encountered while constructing and processing various conceptual structures associated with Web-based informational sources: semantic representations of written and spoken texts' fragments (in other terms, text meaning representations); high-level conceptual descriptions of visual images; knowledge pieces stored in ontologies; the content of messages sent by computer intelligent agents, etc.

How to find a key to solving this problem? We do know that, using NL, we are able to describe various pieces of knowledge, the semantic content of a visual image, the semantic content of a film, etc. That is why it can be conjectured that a key to elaborating a new paradigm of the described kind could be the construction of a broadly applicable and flexible Conceptual Metagrammar. It is to be a collection of the rules (or partial operations) enabling us to construct step by step an SR of practically arbitrary sentence or discourse pertaining to mass spheres of professional activity of people. In [29], the term "a comprehensive semantic formal environment" is used in the same sense.

The prefix "meta" in the term "metagrammar" means that such rules are to use the information associated with the classes of conceptual units. That is why we should be able to employ the same system of rules with different conceptual vocabularies.

4 Theory of K-representations as a source of a broadly applicable and flexible semantic formalism

Happily, a solution to the formulated problem is already available. It is given by the theory of K-representations (TKR). It should be underlined that its approach to describing semantic structure of NL-texts is free from the listed shortcomings of AMR.

In order to better understand the peculiarity of TKR, let's establish an analogy with bionics. Bionics studies the peculiarities of the structure and functioning of the living beings in order to discover the new ways of solving certain technical problems. TKR was developed as a consequence of fulfilling a system analysis of the basic expressive mechanisms of NL and putting forward a conjecture about a system of partial operations on conceptual structures underpinning these expressive mechanisms.

4.1 Two versions of a broadly applicable and flexible conceptual metagrammar

The first basic constituent of TKR is two versions of a mathematical model (Model 1) describing a system of ten partial operations on conceptual structures. The first version (Model 1-A) is published in [17]. It should be noticed that the 9th operation introduced in [17] is modified in [18]. Model 1-A is the kernel of the theory of restricted standard knowledge languages (RSK-languages). The predecessor of this theory is the theory of S-calculuses and S-languages (see [11] and a retrospective outline in Section 2.3 of [25]). The second version (Model 1-B) is published in the monographs [21, 25] and is the kernel of the theory of standard knowledge languages (SK-languages).

Each version of the Model 1 gives us formal means being convenient for building SRs of, likely (it is a hypothesis), arbitrarily complex sentences and discourses in NL pertaining to mass spheres of professional activity (engineering, business, medicine, etc.).

The difference between the Models 1-A and 1-B is as follows. Model 1-A allows us to proceed from only one angle of look at an entity from the considered thematic domain. To the contrary, Model 1-B makes it possible to consider an entity from several possible angles of look.

Example. Both Model 1-A and Model 1-B consider a finite set of symbols *St* and the countable non-intersecting sets of symbols *X* and *V*. The elements of the sets *X* and *V* are interpreted respectively as primary informational units and variables. The set *St* (its elements are called sorts) is a subset of *X*. Suppose also that the Model 1-A includes a mapping *tp1* from the union of *X* and *V* into the countable set of symbols *Types1*, and the Model 1-B includes a mapping *tp2* from the union of *X* and *V* into the countable set of symbols *Types2*. Here *Types1* and *Types2* contain the symbols and strings interpreted as semantic characteristics of entities from the considered domains. Both *Types1* and *Types2* include the subset of sorts *St*, and *Types1* is a subset of *Types2*.

Suppose that *X* includes the unit *D.Mendeleev*, it denotes the famous Russian chemist Dmitry I. Mendeleev, the author of the periodical table of elements. Let *St* include the sorts *ints* and *dyn.phys.ob* ("intelligent system" and "dynamic physical object"). Then it is possible that either *tp1(D.Mendeleev) = ints* or *tp1(D.Mendeleev) = dyn.phys.ob*, but *tp2(D.Mendeleev) = ints * dyn.phys.ob.*

Subsection 4.3 very shortly, without numerous mathematical details, characterizes ten partial operations from Model 1-A and Model 1-B. Due to a very general level of discussion, the material of Subsection 4.3 illustrates the partial operations both from Model 1-A and Model 1-B. Due to the lack of mathematical details, the shortly described operations may seem to be very simple. However, Model 1-A and Model 1-B are strictly mathematical models, they define respectively new classes of formal languages: the classes of RSK-languages and SK-languages. These models were developed due to the invention of an original methodology of constructing inductive definitions of formal objects with complex structure (see [17, 25]).

The analysis of the scientific literature on artificial intelligence theory, mathematical and computational linguistics shows that today the class of SK-languages opens the broadest prospects for building semantic representations (SRs) of NL-texts (i.e., for representing meanings of NL-texts in a formal way).

4.2 The models of linguistic database and algorithms of semantic parsing

The *second basic constituent* of TKR is two broadly applicable mathematical models of a linguistic database (LDB) [21, 25]. The models describe the frames expressing the necessary conditions of the existence of semantic relations, in particular, in the word combinations of the kinds "Verbal form (verb, participle, gerund) + Preposition + Noun", "Verbal form+ Noun", "Noun1 + Preposition + Noun2", "Noun1 + Noun2", "Number designation + Noun", "Attribute + Noun", "Interrogative word + Verb". The expressive power of SK-languages enables us to associate the lexical units with the appropriate simple or compound semantic units. The models describe the logical structure of LDB being the components of NL-interfaces to intelligent databases as well as to other applied computer systems.

The *third basic constituent* of TKR is several complicated, strongly structured algorithms carrying out semantic parsing of texts from some practically interesting sublanguages of NL. The first and second algorithms, called *SemSyn* and *SemSynt1* respectively, are based on the elaborated formal models of LDB. The algorithm *SemSyn* [21] transforms a NL-text in its SR being a K-representation, the algorithm *SemSyn* is described in two final chapters of the monograph [21], and the algorithm *SemSynt1* is set forth in Chapters 9 and 10 of the monograph [25].

An important feature of these algorithms is that they don't construct any syntactic representation of the inputted NL-text but directly find semantic relations between text units. Since numerous lexical units have several meanings, the algorithm uses the information from a linguistic database and linguistic *context* for choosing one meaning of a lexical unit among several possible meanings.

The other distinguished feature is that these structured algorithms are completely described with the help of formal tools, that is why they are problem independent and don't depend on a programming system. The algorithm *SemSyn* is implemented in the programming language PYTHON. Additional information about the algorithms of semantic parsing proposed by TKR can be found in Section 5.

4.3 About ten partial operations on conceptual structures

The expressions of SK-languages will be called below the K-strings. If *Expr* is an expression in NL and a K-string *Semrepr* can be interpreted as an SR of *Expr*, then *Semrepr* will be called a possible K-representation (KR) of the expression *Expr*.

The KRs of NL-texts are formed from the primary informational units, the variables, and several service symbols by means of an iterative process of applying the operations of building well-formed formulas Op[1], …, Op[10]. The initial set of simplest formulas is determined by a special formal object called a conceptual basis (c.b.) and playing the role of the simplest knowledge base [21, 25]. The language determined by the considered c.b. *B* and the operations Op[1], … Op[n] (they are defined by the special statements, or rules, P[1], …, P[10]) is denoted as *Ls(B)* and is called *the standard knowledge language (SK-language) in the basis B* [21, 25].

The rule P[0] provides an initial stock of formulas. For example, if the string *mouse1* is an element of a certain primary informational universe *X*(*B*), then *mouse1* is a formula of *Ls(B)*.

For arbitrary c.b. *B*, let *Degr(B)* be the union of all Cartesian *m*-degrees of *Ls(B),* where *m* is not less than 1. Then the meaning of the rules of constructing well-formed formulas P[1], ..., P[10] can be explained as follows: for each k from 1 to 10, the rule P[k] determines a partial unary operation *Op[k]* on the set *Degr(B)* with the value being an element of *Ls(B).*

Let's consider a short introduction to the partial operations for constructing formal representations of structured meanings Op[1], …, Op[10].

The operation Op[1] can be used to join intentional quantifiers to the designations of the notions and produce the formulas like

certain car, *certain car * (Manufacturer, IBM), all car * (Manufacturer, BMW).*

The operation Op[2] can be used to construct the formulas like $f(a_1, \dots, a_n)$, where *f* is a functional symbol, and $a_1, \dots, a_n$ are the well-formed formulas of *Ls(B).* For example, *Area (certain country)* is a well-formed formula of a certain SK-language *Ls(B)*. The operation Op[3] can be used to construct the expressions of the form *(a ≡ b)*. E.g., *(Area (certain country) ≡ x12).*

The operation Op[4] can be used to construct the formulas like $rel\ (a_1, \dots, a_n)$, where *rel* is a relational symbol, and $a_1, \dots, a_n$ are the formulas of *Ls(B)*. E.g., *Less(Area (certain country), 600,000/sq.km).*

The operation Op[5] allows us to mark KRs by some variables from the set of variables. For example, if a part of a KR looks like *certain file1 * (Extension, ".docx") : v1*, then we can refer to the expression *certain file1 * (Extension, ".docx")* in another part of a K-representation, using *v1*.

The operation Op[6] provides the possibility to construct K-representations in the form ¬*Formula*, for example ¬ *car*. The operation Op[7] allows us to use conjunction and disjunction in the formulas, e.g., (*airplane* ∨ *helicopter*), (*mathematician* ∧ *painter*).

The operation Op[8] can be used to build compound designations of the notions in the form

$$concept * (r_1, value_1) \dots (r_n, value_n),$$

where *concept* is an element of a primary informational universe *X(B)* denoting a notion, $r_1, \dots r_n$ are the names of functions or relations, and the $value_1, \dots, value_n$ are well-constructed formulas. This operation allows us to construct the formula *country *(Location, Europe)(Capital, Vienna)* being a KR of the expression "a country in Europe with the capital Vienna".

The operation Op[9] allows us to use quantifiers ∀ and ∃ like in FOL. The operation Op[10] enables us to build the representations of ordered *n*-tuples as the expressions of the form $<a_1, \dots a_n>$, where $a_1, \dots a_n$ are some well-constructed formulas. E.g., this operation can be used to construct the KR *<Place, Backup-drive>, <Time, Midnight>, <Frequency, Everyday>*.

These *n*-tuples could be used to construct representations of complex verb constructions. For example: *Delete(<Object, all file1*(Size, 0)(Extension, ".txt")>, <Time, Midnight>, <Frequency, Everyday>).*

4.4 SK-languages as a tool of describing semantic structure of sentences

Before to consider below a number of examples illustrating a correspondence between an expression in NL and its possible KR, let's agree that the string *Semrepr* is to be interpreted as a possible KR of the regarded expression in NL.

Compound semantic descriptions of objects and sets of objects. The key role is played by the interaction of the operations Op[8], Op[1], and Op[5]. Using the operation Op[8] at the last step of constructing a formula and any of the operations Op[1], …, Op[10] at the previous steps, it is possible to construct an expression of the form $conc * (rel_1, d_1) \dots (rel_n, d_n)$, where *conc* is a simple (non-structured) designation of a notion, $n \geq 1$, for $k = 1, \dots, n$ rel_k either is a name of the function with one argument or the name of a binary relation. In the first case d_k designates the value of the function rel_k and in the second case d_k designates the second attribute of the relation rel_k .

Applying consequently the operations Op[1] and Op[5], we can obtain an expression of the form

qtr conc * *(*rel_1, d_1*)…(*rel_n, d_n*): var,*

where *qtr* is an intensional quantifier (in particular, it may correspond to the meanings of the words and expressions "a certain", "any" "all"), *var* is a variable.

Example 1. We can construct compound designations of the entities mentioned in texts. For example, the expression "a French textbook on biology" may be associated with the semantic image

*certain textbook1 *(Country, France)(Activity-field, biology) : x15.*

Example 2. It is possible to build compound designations of the mentioned sets, e.g., *certain set1 * (Number-of-elements, 4)(Qualitative-composition, container1 * (Content1, ceramics * (Country-producer, (India OR China)))) : S7,* where *set1* designates the notion "finite set".

Building semantic representations of compound infinitive constructions.

Example 1. Let Goal1 = "To receive a M.Sci. degree in business informatics at the Higher School of Economics (Moscow) and to found a company on e-business". Then a possible K-representation of Goal1:

*(receiving1 * (Institution-role, certain university * (Name1, "Higher School of Economics")(Location, certain city * (Name1, "Moscow") : x1))*
*(Document-role, certain acad-degree **
(Kind, M.Sci.)(Field1, business-informatics) : x2)
*∧ founding1 * (Organization-role, certain*
*firm1 * (Field1, e-business) : x3))*

Representation of the meanings of sentences with indirect speech. Let T1 = "When Mr. Peter Smith announced that he would visit Montpelier in April?". Then *Semrepr = Question (t1, Situation (e1, informing1 * (Time, certain mom * (Before, #now#) : t1) (Agent1, certain man * (First-name, "Peter")(Surname, "Smith") : x1)(Inform-content, Situation (e2, visit1 * (Agent1, x1)(Location2, certain city * (Name1, "Montpellier") : x2)(Time, Nearest-month-future (April, #now#)))))).*

Representing the meanings of sentences with subordinate clauses of purpose. Let T2 = "Mr. Peter Smith, a Vice-President of the firm "Rainbow", announced yesterday that he would visit Montpelier in April in order to sign an agreement with the company "CIRAD". Then

*Semrepr = Situation (e1, informing1 * (Time, Previous-day (#now#)) (Agent1, certain man * (First-name, "Peter")(Surname, "Smith") : x1)(Inform-content, Situation (e2, visit1 * (Agent1, x1)(Location2, certain city * (Name1, "Montpellier") : x2)(Time, Nearest-month-future(April, #now#))(Goal, signing2 * (Inform-object, certain agreement1 : x3)(Business-partner, certain company1 * (Name1, "CIRAD") : x4))))).*

Semantic representation of the homogeneous members of sentence. Let T3 = "Jean would ike to visit during this summer either Vienna, Bratislava, and Prague or Bergen, Oslo, and Stockholm". Then

*Semrepr = Situation (e1, intention * (Time, #now#) (Emotional-agent, certain man * (First-name, "Jean) : x1)(Goal, visit1 * (Time, Nearest-season(summer, #now#))(Location2, ((certain city * (Name1, "Vienna" : x2) ∧ certain city * (Name1, "Bratislava" : x3) ∧ certain city * (Name1, "Prague" : x4) ∨ (certain city * (Name1, "Bergen" : x5) ∧ certain city * (Name1, "Oslo" : x6) ∧ certain city * (Name1, "Stockholm" : x7)))))).*

Semantic descriptions of the expressions with the words "a notion", "a term". Let S1 = "The term gene was first coined in 1909 by a Danish botanist, Johannsen, and was derived from the term pangen introduced by De Vries. Then

*Semrepr1 = Situation (e1, introduction1 * (Notion-name, certain notion * (called, "gene") : c1)(Agent1, certain botanist1* (Surname, "Johannsen")(Country-role, Denmark) : x1)(Time, 1909)) ∧ Situation (e2, derivation1 * (Notion-name, c1)(Agent1, x1)(Source-notion, certain notion * (Called, "pangen") (Authorship, certain person * (Surname, "De Vries") : x2).*

4.5 SK-languages as a tool of describing semantic structure of discourses and representing knowledge pieces

Example 1. Let Disc = S1. S2, where S2 = "This information is given in the textbook "Emery's Elements of Medical Genetics" by D. Turnpenny and S. Ellard, its 12th edition was published by Elsevier in 2005". Then Disc may have a KR of the form

(Semrepr1 : P1 ∧ Information-source (P1, Semrepr2)), where *Semrepr2* is the following possible KR of the sentence S2 :

*certain textbook1 * (Title, "Emery's Elements of Medical Genetics")(Authorship, (D. Turnpenny ∧ S. Ellard))(Edition-number, 12)(Publishing-house, Elsevier)(Year, 2005) : x3.*

Here *P1* is the variable marking the meaning of the first phrase of the text Disc.

Example 2. Let Def = "Control gene is a gene which can turn other genes on or off". Then

*Semrepr3 = (Control-gene ≡ gene * (Is-able, (turning-on * (Object-bio, some gene : Set1) ∧ turning-off * (Object-bio, Set1)))).*

Example 3. It is possible to construct a different KR of the definition Def, it will reflect the metadata of information piece, indicating the edition, the authors, and year of publication. In this case

*Semrepr-with-metadata = certain inform-object * (Content1, Semrepr3)(Authorship, (D.Turnpenny ∧ S.Ellard))(Publishing-house, Elsevier)(Year, 2005)*
(Title, "Emery's Elements of Medical Genetics")(Edition-number, 12).

5 Principal distinctive features of two original approaches to semantic parsing

The theory of K-representations not only introduced a new class of formal languages (the class of SK-languages) for building SRs of complex sentences and discourses. It also used the definition of this class of formal languages as a starting point for developing two broadly applicable mathematical models of a linguistic database ([22], Chapter 6 of [21], and Chapter 7 of [25])) and an original method of extracting structured meanings from NL-texts (Chapter 8 of [25]). We use this term for denoting a method of developing the multilingual algorithms of semantic-syntactic analysis of texts in NL. Such algorithms transform the texts from certain sublanguages of NL into SRs (in other terms, text meaning representations). For building SRs, the class of SK-languages is used. The input texts may be at least from broad and practically interesting sublanguages of English, German, and Russian languages.

The proposed method underpinned the development of a multilingual algorithm of semantic parsing *SemSynt1* (Chapters 9 and 10 of [25]). It is the composition of two algorithms called *BuildMatr1* and *BuildSem1*. The algorithm *BuildMatr1* can be qualified as *an original algorithm of semantic role labeling*. The input texts may be the questions of many kinds, the commands, the sentences, and the discourses. The output of *BuildMatr1* (more exactly, its principal part) is a special string-digital matrix *Matr* called *a matrix semantic-syntactic representation (MSSR)* of the input text. The matrix *Matr* is dynamically linked with an auxiliary data structure being a two-dimensional array *Arls*. In case an elementary meaningful text unit (or a token) *wd* has *N* different meanings, the array Arls will include *N* consequent rows,

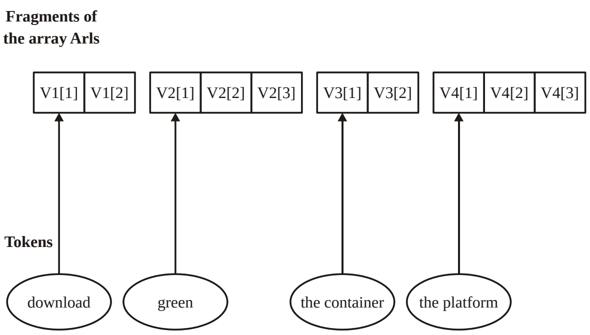

Figure 1: Initial graph and mapping determined by an MSSR Matr.

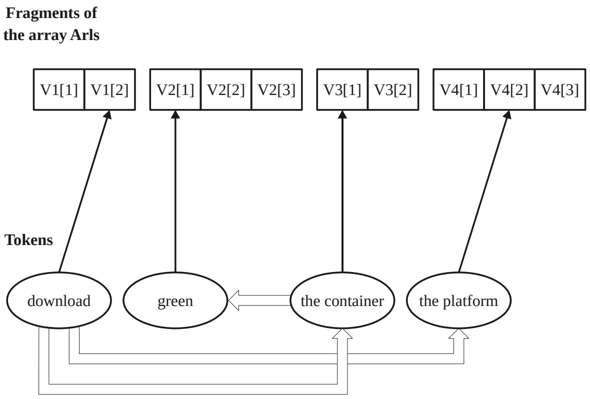

Figure 2: Final graph and mapping determined by an MSSR Matr.

where for *k* = 1, …, *N* the *N*-th row stores the information associated with the *k*-th meaning of *wd*.

The configuration of an MSSR *Matr* changes during semantic-syntactic processing of the input text. Each configuration determines, in particular, a marked oriented graph with the vertices being the distinguished elementary meaningful text units (or tokens) and a mapping from the subset of the vertices of this graph corresponding to lexical items to the set of meanings (or values) associated with these lexical items via the array *Arls*. Before the start of text's processing, an edge from each lexical unit *wd* goes to the first row of *Arls* (that is, the row with the minimal ordered number) storing the semantic units associated with *wd*.

Figure 1 illustrates this situation for processing the command "Download the green container on the platform". Here V1[1] is the value *downloading1* (downloading a file), V1[2] is the value *downloading2* (downloading a transportable physical object); V2[1] is the value *green-colour*, V2[2] is the value *not-ripe*, V2[3] designates the value *a-member-of-green-movement*; V3[1] is the value *thing-container*; V3[2] is the value *data-structure-of-RDF*; V4[1] is the value *computer-platform*, V4[2] is the value *railway-station-platform*, V4[3] is the value *political-platform*. Figure 2 illustrates the final situation.

The output of the algorithm *BuildMatr1* is the input of the algorithm *BuildSem1*. It transforms the information represented by an MSSR *Matr* of the input text into its possible SR. It is a KR of the input text.

Example. The command "Download the green container on the platform" can be associated with its possible KR of the form

*Command (#Operator#, #Executor#, #now#, downloading2 * (Object1, certain thing-container * (Colour, green) : x1)(Destination, certain railway-station-platform : x2)).*

The paper [44] expands the method introduced in Chapter 6 of [25]. On the one hand, the input language of the algorithm *BuildMatr1* is enriched by means of the phrases expressing (a) the values of functions, (b) the restrictions of the functions' values, (c) the relations between various objects formed with the help of comparative adjectives.

On the other hand, it is well known that many notions corresponding to the words and word combinations from NL-texts are too general in order to be used for the interaction with a database. For instance, these are the concepts "IT-specialist" and "alumni". That is why it is proposed to use for semantic parsing of NL-texts not only a linguistic database but also a linguistic knowledge base (LKB). It may consist of the K-strings of the form illustrated by the following example:

*(IT-specialist ≡ person * (Qualification, (programmer ∨ database-administrator ∨ web-programmer)).*

Let's call *unfolding concepts* the concepts being the left parts of some expressions in the LKB. The proposed final step of processing NL-texts is to replace all semantic items from the constructed primary SR belonging to the subclass of unfolding concepts by the less general concepts with the help of the definitions stored by the used LKB (it may

be interpreted as a part of ontology). E.g., the concept “IT-specialist” will be replaced by the compound concept
*person * (Qualification, (programmer ∨ database-administrator ∨ web-programmer)).*

The paper [45] introduces a highly compact way of describing formal structure of linguistic databases (semantic-syntactic component) and of presenting the algorithms of semantic parsing. The paper contains the algorithm of semantic parsing SemSyntRA, developed under the framework of the proposed approach (see also next section).

6 Applications of the K-representations theory

The arguments stated above and numerous additional arguments set forth in the monograph [25] give serious grounds to conclude that the class of SK-languages, provided by TKR, can be interpreted as the first comprehensive semantic formal environment for studying various semantics-associated problems of developing an MSW.

It seems to be reasonable to say about two levels of applying TKR to solving practical tasks. The *first level* is the direct use in the design of NL processing systems of a mathematical model of a linguistic database introduced in Chapter 7 of the monograph [25] and of the algorithm of semantic parsing *SemSynt1* described in Chapters 9 and 10 of the same monograph. This algorithm is multilingual: its input texts may be the questions of many kinds, statements, and commands from the sublanguages of English, German, and Russian languages. The mentioned model and algorithm were applied by the author and his Ph.D. students to the design of an NL-interface of a recommender system [41], to the design of an advanced semantic search system [30], and to the design of an NL-interface of an applied intelligent system making easier the interaction of a user with the file system of a computer [44, 45]. Two versions of this system are called NLC-1 [44] and NLC-2 [45] (here NLC = Natural Language Commander).

Example. Let’s look how NLC-1 processed the following user instruction: “Copy music files from “Download” folder to folder with name “Music” or “My music” on backup drive if their size is less than 1 GB”. This instruction has the following primary K-representation constructed by *SemSynt2* – a modification of the algorithm *SemSynt1*:

If-then(Less(SizeOf(all music1(Place1, certain folder1 * (Name1, "Download")):o1), 1/GB), Command (#Operator#, #Executor#, #now#, copying*(Source1, o1)(Destination1, certain folder1 * (Name1, ("Music" ∨ "My music"))(Place1, certain backup-drive)))).*

Now if knowledge base of NLC-1 contains the K-strings *(music1 ≡ file1 * (Extension, ("mp3" ∨ "ogg" ∨ "wav" ∨ “aac")), (backup-drive ≡ drive1 *(Name1,"F"))* and knowledge management system includes the rule (*x*, (*x* ≡ *y*) ├ *y*), then NLC-1 transforms the constructed primary KR of user instruction into the secondary KR

If-then(Less(SizeOf(all file1(Extention, ("mp3" ∨ "ogg" ∨ "wav" ∨ "aac"))(Place1, certain folder1 * (Name1, "Download")) : o1), 1/GB), Command((#Operator#, #Executor#, #now#, copying*(Source1,o1)(Destination1, certain folder1 *(Name1, ("Music" ∨ "My music"))(Place1, certain drive1 * (Name1, "F"))))).*

Then the result shell script is

if [$(du -cb "Download/*.mp3" "Download/*.ogg" "Download/*.wav" "Download/*.acc"|grep total|sed -e "s/\s.*$//g") -le 1000000000]; then cp "Download/.mp3" "Download/*.ogg" "Download/*.wav" "Download/*.acc" $(ls /f/|grep -iE "^Music$|^My music$" / head -n1); fi.*

Written in Haskell programming language, NLC-1 is a flexible and scalable application. It can be configured by a researcher for different domains and underlying shells. The paper [45] describes a modified theoretical foundation of the second version NLC-2.

The great advantages of the proposed comprehensive semantic formal environment are promised by the *second level* of applications: it is the case of using SK-languages for describing lexical semantics, representing semantic content of sentences and discourses in NL, building models of advanced ontologies, constructing semantic annotations of Web-documents (see Section 6.2 of [25]), and forming high-level conceptual descriptions of visual images (see Section 6.3 of [25]) in numerous scientific centres and research groups throughout the world.

7 A contribution to developing a Multilingual Semantic Web

The process of endowing the existing Web with the ability of understanding many natural languages is an objective ongoing process. The analysis has shown that there is a way to increase the total successfulness, effectiveness of this global decentralized process. It would be especially important with respect to the need of cross-language conceptual information retrieval and question - answering. The way proposed in [25-29] is a possible new paradigm for the mainly decentralized process of endowing the existing Web with the ability of processing many natural languages.

The principal idea of a new paradigm is as follows. There is *a common thing* for the various texts in different natural languages. This common thing is the fact that *the NL-texts have the meanings.* The meanings are associated not only with NL-texts but also with the visual images (stored in multimedia databases) and with the pieces of knowledge from the ontologies.

That is why the great advantages are promised by the realization of the situation when a unified formal semantic environment is being used in different projects throughout the world for reflecting structured meanings of the texts in various natural languages, for representing knowledge about application domains, for constructing semantic annotations of informational sources and for building high-level conceptual descriptions of visual images.

The analysis of the expressive power of SK-languages (see the chapters 3 – 6 of [25] shows that the SK-languages can be used as a unified formal semantic environment of the kind. This idea underlies an original strategy of transforming step by step the existing Web into a Semantic Web of a new generation, where its principal distinguished feature would be the well-developed ability of NL processing; it can be also qualified as a Multilingual Semantic Web. The versions of this strategy are published in [25-29].

8 Conclusion

Computational semantics has received a firm theoretical ground. The SK-languages, introduced by the theory of K-representations, open new prospects for formalizing lexical semantics, representing semantic content of sentences and discourses in NL, building models of advanced ontologies, forming high-level conceptual descriptions of visual images, and constructing semantic annotations of Web-documents in numerous scientific centres and research groups. Many existing projects on NL processing including semantic parsing have received an appropriate theoretical framework for next stages of research. For an MSW it is also very important that SK-languages provide a convenient intermediary level for moving from NL input to OWL-based ontologies.

This paper provides additional arguments in favour of the conjecture formulated in [24-29]: TKR can be and should be used as a comprehensive and flexible basic formal tool for solving the tasks of developing an MSW associated with semantics of NL.

References

[1] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N. (2013). Abstract Meaning Representation for Sembanking. *In: Proceedings of the 7th ACL Linguistic Annotation Workshop and Interoperability with Discourse*, Sofia, Bulgaria, August 8-9, 2013 (2013)(www.aclweb.org/anthology/W13-2322; retrieved *2016-03-12*)

[2] Banarescu, L., Bonial, C., Cai, S., Georgescu, M., Griffitt, K., Hermjakob, U., Knight, K., Koehn, P., Palmer, M., Schneider, N. (2015). *Abstract Meaning Representation (AMR) 1.2.2 Specification*; github.com/amrisi/amr-guidelines/blob/master/amr.md.

[3] Blanco E., Moldovan D. (2014). Leveraging verb-argument structures to infer semantic relations. *In Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*, Gothenburg, Sweden, April 26-30, 2014. ACL, pp. 145-154.

[4] Bordes, A., Glorot, X., Westion, J., Bengio, Y. (2012). Joint learning of words and meaning representations for open-text semantic parsing. *Proc. of the 15th Intern. Conf. on Artificial Intelligence and Statistics (AISTATS) 2012*, LasPalmas, Canary Islands. Vol. 22, pp. 127-135.

[5] Buitelaar P., Choi K.-S. Cimiano P., Hovy E. H. (Eds.)(2012). Report from Dagstuhl Seminar 12362 “The Multilingual Semantic Web” (2 – 9 September, 2012). Schloss Dagstuhl: Leibniz-Zentrum fuer Informatik.

[6] Cambridge Semantics Inc., The Smart Data Company, Web-page (retrieved 14.10.2016).; http://www.cambridgesemantics.com/semantic-university/nlp-and-semantic-web#..

[7] Cimiano, P., Haase, P. et al. (2008). Towards Portable Natural Language Interfaces to Knowledge Bases – the Case of the ORAKEL System. *Data and nowledge Engineering*, Vol. 65, No. 2., pp. 325-354.

[8] Clark P., Harrison P. (2008). Boeing’s NLP System and the Challenges of Semantic Representation. In: Proc. SIGSEM Symposium on Text Processing (STEP’08), Venice, Italy, ACL, pp. 263-276.

[9] Das, D., Chen, D., Martins, A. F. T., Schneider, N., Smith, N. A. (2014). Frame-Semantic Parsing. *Computational Linguistics*, Vol. 40, No. 1, pp. 9-56.

[10] Fillmore C., Johnson C. R., Petruck M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, Vol. 16, No. 3, pp. 235-250.

[11] Fomitchov, V. A. (1984). Formal systems for natural language man-machine interaction modelling. Artificial Intelligence. *Proc. of the IFAC Symposium. Leningrad, USSR, 4-6 Oct. 1983, Ponomaryov, V.M. (ed.)*, Oxford, UK, Pergamon Press Ltd., New York, Pergamon Press Inc., 1984, pp. 203-207 (IFAC Proc. Series, 1984, No. 9).

[12] Fomichov, V. A. (1988) *Representing Information by Means of K-calculuses. Textbook*. Moscow, The Moscow Institute of Electronic Engineering, 1988.

[13] Fomichov, V. A. (1992). Mathematical models of natural-language-processing systems as cybernetic models of a new kind. *Cybernetica. Quarterly Review of the International Association for Cybernetics* (Belgium, Namur).,Vol. 35, No. 1, pp 63–91.

[14] Fomichov, V. A. (1993). Towards a mathematical theory of natural language communication. *Informatica. An Intern. Journal of Computing and Informatics (Slovenia)*, ol. 17, No. 1, pp. 21–34.

[15] Fomichov, V. A. (1993). K-calculuses and K-languages as powerful formal means to design intelligent systems processing medical texts. *Cybernetica (Belgium)*, 993, Vol. XXXVI, No. 2, pp.161-182.

[16] Fomichov, V. A. (1994). Integral Formal Semantics and the design of legal full-text databases. Cybernetica (Belgium), ol. XXXVII, No. 2, pp. 145-177.

[17] Fomichov, V. A. (1996). A mathematical model for describing structured items of conceptual level.

Informatica. An International Journal of Computing and Informatics (Slovenia), Vol. 20, No. 1. pp. 5–32.

[18] Fomichov V. A. (1998). Theory of restricted K-calculuses as a comprehensive framework for constructing agent communication languages. Fomichov, V.A., Zeleznikar, A.P. (eds.), Special Issue on NLP and Multi-Agent Systems. Informatica. An Intern. Journal of Computing and Informatics (Slovenia), Vol. 22, No. 4, pp. 451-463.

[19] **F**omichov, V. A. (2000). An ontological mathematical framework for electronic commerce and semantically-structured Web. Zhang, Y., Fomichov, V.A., Zeleznikar, A.P. (Eds.) Special Issue on Database, Web, and Cooperative Systems. Informatica. An Intern. Journal of Computing and Informatics (Slovenia), Vol. 24, No. 1, pp. 39-49.

[20] Fomichov, V. A. (2002). Theory of K-calculuses as a powerful and flexible mathematical framework for building ontologies and designing natural language-processing systems. Andreasen, T., Motro, A., Christiansen, H., Larsen, H.L. (Eds.), Flexible Query Answering Systems, 5th Intern. Conference, FQAS 2002, Proceedings, Lecture Notes in Artificial Intelligence, Vol. 2522, Springer: Berlin, Heidelberg, New York, pp. 183-196.

[21] Fomichov, V.A. (2005a). The Formalization of Designing Natural Language Processing Systems. Moscow: MAX Press, 368 p. (in Russian).

[22] Fomichov, V.A. (2005b). A new method of transforming natural language texts into semantic representations. *Informational Technologies*, Moscow, No. 10, pp. 25-35 (in Russian).

[23] Fomichov V. A. (2007). *Mathematical foundations of representing the content of messages sent by computer intelligent agents*. Moscow, State University – Higher School of Economics, Publishing Hiuse "TEIS", 2007 – 176 p. (in Russian)

[24] Fomichov, V.A. (2008). A comprehensive mathematical framework for bridging a gap between two approaches to creating a Meaning-Understanding Web. *Intern. Journal of Intelligent Computing and Cybernetics*,. Vol. 1, No. 1, pp. 143-163.

[25] Fomichov, V. A. (2010a) Semantics-Oriented Natural Language Processing: Mathematical Models and Algorithms, Springer, New York, Dordrecht, Heidelberg, London. - 352 p.

[26] Fomichov, V. A. (2010b) Theory of K-representations as a comprehensive formal framework for developing a Multilingual Semantic Web. *Informatica. An Intern. Journal of Computing and Informatics (Slovenia)*, Vol. 34, No. 3, pp.. 387-396.

[27] Fomichov, V. A. (2011). The prospects revealed by the theory of K-representations for bioinformatics and Semantic Web. *Actes de la 18e conference sur le Traitement Automatique des Langues Naturels. Actes de la 15e Rencontre des Etudiants Cercheurs en Informatique pour le Traitement Automatique des Langues. France, Montpellie, 27th June - 1st July 2011* Vol. 1: Actes: articles longs. Montpellier : AVL Diffusion, pp 5-20.

[28] Fomichov, V.A. (2013). A broadly applicable and flexible conceptual metagrammar as a basic tool for developing a Multilingual Semantic Web. In: Metais E., Meziane F., Saraee M., Sugumaran V., Vadera S. (Eds.) Natural Language Processing and Information Systems. 18th Intern. Conference on Applications of Natural Language to Information Systems, NLDB 2013. Salford, UK, June 2013, Proceedings. Lecture Notes in Computer Science, Vol. 7934, Springer, Berlin, Heidelberg, pp. 249-259.

[29] Fomichov, V. A. (2014) SK-languages as a comprehensive formal environment for developing a Multilingual Semantic Web. Decker H., Lhotská L., Link S., Spies M., Wagner R.R. (Eds.). Database and Expert Systems Applications, 25th Intern. Conference, DEXA 2014, Munich, Germany, September 1-4, 2014, Part I, Proceedings. Lecture Notes in Computer Science, Vol. 8644, Cham: Springer International Publishing Switzerland, pp. 394-401.

[30] Fomichov, V.A., Kirillov, A.V. (2012). A formal model for constructing semantic expansions of the search requests about the achievements and failures. *Artificial Intelligence: Methodology, Systems, and Applications, Ramsay A., Agre G. (Eds.), Lecture Notes in Computer Science*, Vol. 7557, Springer, Berlin, Heidelberg, pp. 296–304.

[31] Gildea, D., Jurafsky, D. (2002). Automatic Labeling of Semantic Roles. *Computational Linguistics*, Vol. 28, No. 3, pp. 245 - 288.

[32] Google Hummingbird (2016); https://en.wikipedia.org/wiki/ Google_Hummingbird (retrieved 10.11.2016).

[33] Kingsbury, P., Palmer, M. (2002). From TreeBank to PropBank . *Proceedings. LREC*, 2002.

[34] Langkilde, I, Knight, K. (1998). Generation that exploits corpus-based statistical knowledge. *Proc. of the 36th Annual Meeting of the ACL and 17th International Conference on Computational Linguistics*, Montreal, pp. 704-710.

[35] Li, B., Wen, Y., Bu, L., Qu, W., Xue, N. (2016). Annotating the Little Prince with Chinese AMRs. *Proc. of LAW X – the 10th Linguistic Annotation Workshop.* Berlin, Germany, August 11, 2016, ACL, pp. 7-15.

[36] Liang, P. (2016). Learning executable semantic parsers for natural language understanding. *Communications of the ACM*, Vol. 59, No. 9, pp. 68-76.

[37] Lu, C., Xu, Y., Geva, S. (2008). Web-based query translation for English-Chinese CLIR. *Computational Linguistics and Chinese Language Processing (CLCLP)*, pp. 61-90.

[38] Marcus M. P., Marcinkiewicz M. A., Santorini B. (1993). Building a large annotated corpus of English: the Penn Treebank. *Computational Linguistics*, Vol. 19, No. 2.

[39] Marquez L., Carreras X., Litkowski K. C., Stevenson S. (2008) Semantic Role Labeling: an Introduction to

the Special Issue. *Computational Linguistics*, Vol. 34, No. 2, pp. 145-159.

[40] Ontos GmbH company Web-page (2016): www.ontos.com.

[41] Pravikov, A.A., Fomichov, V.A. (2010). Development of a recommender system with a natural language interface on the basis of semantic objects' mathematical models. *Business Informatics. Interdisciplinary scientific-practical journal*, Moscow, State University – Higher School of Economics, 2010, No. 4 (14), pp. 3-11.

[42] Punyakanok V., Roth D., Yih W. T. (2008). The importance of syntactic parsing and inferencing in semantic role labeling. *Computational Linguistics*, Vol. 34, No. 2, pp.. 257-287.

[43] Pust, M., Hermjakob, U., Knight, K. , Marcu, D., May, J. (2015). Parsing English into abstract meaning representation using syntax-based machine translation. In *Proc. of the EMNLP 2015*, Lisbon, pp. 1143-1154.

[44] Razorenov, A. A., Fomichov, V. A. (2014). The Design of a Natural Language Interface for File System Operations on the Basis of a Structured Meanings Model. In *Procedia Computer Science*, Elsevier. V. 31. P. 1005-1011; open access, URL: http://authors.elsevier.com/sd/article/S1877050914005304.

[45] Razorenov, A. A., Fomichov, V. A. (2016). A new formal approach to semantic parsing of instructions and to file manager design. In *Database and Expert Systems Applications, 27th Intern. Conference, DEXA 2016, Porto, Portugal, September 5-8, 2016, Part I, Proceedings.* Lecture Notes in Computer Science. V. 9827. Cham: Springer International Publishing Switzerland, pp. 416-430.

[46] Rindflesh, T.C., Kilicoglu, H., Fiszman, M., Roszemblat, G., Shin, D. (2011). Semantic MEDLINE: An Advanced Information Management Application for Biomedicine. Information Services and Use, IOS Press Vol. 1, pp. 15-21.

[47] Sawai, Y., Shindo, H., Matsumoto, Y. (2015). Semantic structure analysis of noun phrases using abstract meaning representation. *Proc. of the 53rd Annual Meeting of the ACL (Volume 2: Short papers)*, Beijing, pp. 851-856.

[48] Schubert, L.K., Hwang, C.H. (2000). Episodic Logic meets little red riding hood: A comprehensive, natural representation for language understanding. In *Iwanska, L., Shapiro, S.C. (eds.), Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, MIT/AAAI Press, Menlo Park, CA, and Cambridge, MA, pp. 111-174.

[49] Stellato, A. (2016). A language-aware Web will give us a bigger and better Semantic Web. MSW 2015. Multilingual Semantic Web. Proc. of the Fourth Workshop on the Multilingual Semantic Web (MSW4) co-located with 12th Extended Semantic Web Conference (ESWC 2015), Portoroz, Slovenia, June 1, 2015, pp. 1-14.

[50] Sullivan, D. (2013). FAQ: All about the new Google "Hummingbird" algorithm. Search Engine Land, 26 September 2013, http://searchengineland.com/google-hummingbird-172816 (retrieved 10.11.2016).

[51] Uchida, H., Zhu, M., Della Senta, T. (1999). A Gift for a Millennium..

[52] Uren, V.S., Lei, Y., Motta, E.: (2008). SemSearch: Refining Semantic Search. In: *Bechhofer, S., Hauswirth, M., Hoffman, J., Koubarakis, M. (Eds.), ESWS 2008, LNCS*, vol. 5021. Springer, Heidelberg, pp. 874-878.

[53] Vogt, M. (2016). How Natural Language Processing will change the Semantic Web. *Semantics, April 13, 2016;* https://2016.semantics.cc/how-natural-language-processing-will-change-semantic-web (retrieved 12.09.2016).

[54] Wang, C., Xue, N., Pradhan, S. (2015). Boosting transition-based AMR parsing with refined actions and auxiliary analyzers. *In Proc. of the 53rd Annual Meeting of the ACL (Volume 2: Short papers)*, Beijing, pp. 857-862.

[55] Werling, K., Gabor, A., Manning, C.D. (2015). Robust subgraph generation improves abstract meaning representation parsing. In *Proc. of the 53rd Annual Meeting of the ACL (Volume 1: Long Papers)*, Beijing, pp. 982-991.

[56] Wilks, Y., Brewster, C. (2006). Natural Language Processing as a Foundation of the Semantic Web. Foundations and Trends in Web Science, Vol. 1, No. 3. Hanover, MA; Delft: now Publ. Inc.

[57] Wintner S. (2009). What science underlies natural language engineering? *Computational Linguistics*, Vol. 35, No. 4, pp. 641-644.

[58] Yi ,S, Loper., E., Palmer, M. (2007) .Can semantic roles generalize across genres? *Proceedings of the Human Language Technologies Conference of the North American Chapter of the ACL.* Rochester, NY, pp. 548-555.