A comprehensive mathematical framework for bridging a gap between two approaches to creating a meaning-understanding web

Received 1 November 2007
Revised 8 December 2007
Accepted 17 December 2007

Vladimir Fomichov
Department of Innovations and Business in the Sphere of Informational Technologies, Faculty of Business Informatics, Higher School of Economics, State University, Moscow, Russia

Abstract

Purpose – Since, the middle of the 1990s, UNO has been funding a large-scale project aimed at designing a family of natural language (NL) processing systems transforming the sentences in various NLs into the expressions of a new language-intermediary called the Universal Networking Language (UNL) and vice versa. The purpose of the paper is to propose a constructive way of developing a semantic networking language (SNL) of a new generation and, as a consequence, to bridge a gap between UNL-based studies and Semantic Web projects.

Design/methodology/approach – The methodological basis of the paper is a new theory of designing semantic-syntactic analyzers of NL texts elaborated by the author of the paper and called the theory of K-representations (knowledge representations). One of its basic components is a mathematical model describing a system of ten partial operations on conceptual structures and determining a new class of formal languages called restricted standard knowledge languages (RSK-languages).

Findings – It is shown that the expressive possibilities of RSK-languages surpass the expressive possibilities of UNL from the standpoint of representing the meanings of discourses, compound goals, descriptions of sets, definitions of notions. It is proposed to use the definition of the class of RSK-languages as a model of a SNL of a new generation in comparison with UNL.

Practical implications – It is also proposed to use the definition of RSK-languages for building semantic annotations of arbitrary web-documents and web-services.

Originality/value – The paper describes an original approach to representing conceptual structure of NL texts.

Keywords Text retrieval, Mathematics, Semantics, Networking, Worldwide web, Programming languages

Paper type Technical paper

1. Introduction

Since, the second half of the 1990s, one has been able to observe the progress of two parallel approaches to adding to the existing web the ability to understand the meanings of electronic documents. On one hand, it has been the activity of the research laboratories of the world wide web consortium and a number of other research centers in the world aimed at developing a Semantic Web. Though officially the task of creating Semantic Web was announced in the beginning of 2001, the possibility to pose

International Journal of Intelligent Computing and Cybernetics
Vol. 1 No. 1, 2008
pp. 143-163

1756-378X
DOI 10.1108/17563780810857176

this task was created by a number of preliminary studies which resulted, in particular, in the development of resource description framework (RDF) – a language for describing the metadata about informational sources and RDF schema specification language (RDF SSL). In the first decade of this century, RDF and RDF SSL became the basis for the development of DAML + OIL and its successor OWL – two languages destined for constructing ontologies. On the other hand, the following fundamental problem has emerged in the middle of the 1990s: how to eliminate the language barrier between the end-users of the internet in different countries. To solve this problem, Uchida *et al.* (1999) proposed a new language-intermediary, using the words of English language for designating informational units and several special symbols. This language, called the Universal Networking Language (UNL), is based on the idea of representing the meanings of separate sentences by means of binary relations. The second motive for the elaboration of UNL was an attempt to create the language means allowing for representing in one format the various pieces of knowledge accumulated by the mankind and, as a consequence, to create objective preconditions for sharing these pieces of knowledge by various computer systems throughout the world.

Since, 1996, UNO has been funding a large-scale project aimed at the design of a family of natural language processing systems (NLPSs) transforming the sentences in various NLs into the expressions of UNL and also transforming the UNL-expressions into sentences in various NLs. For several years, the coordinator of this project was the UNO Institute for Advanced Studies, Tokyo University. At the moment, under the framework of this project, the NLPSs for six official UNO languages are being elaborated (English, Arabic, Spanish, Chinese, Russian, and French), and also for nine other languages, including Japanese, Italian, and German. Since, the beginning of the 2000s, the studies in this direction have been coordinated by the Universal Networking Digital Language Foundation.

The initially scheduled duration of the UNL project, started in 1996, was ten years. That is why it is just the time to analyze the achieved results and to take the right decisions concerning the further studies in this direction. Continuing the line of Fomichov (2004, 2007) with respect to the online monographs (UNL, 2005, 2006), this paper shows that the expressive possibilities of UNL are rather restricted. That is why it is proposed to interpret the language UNL (despite of the linguistic meaning of its title) as a semantic networking language (SNL) of the first generation.

The purpose of this paper is to draw the attention of the designers of semantic informational technologies, first of all, the researchers in the field of Semantic Web, to the theory of knowledge representations (K-representations) (elaborated by the author of this paper) as a comprehensive framework for the creation of a SNL of the next generation possessing much broader expressive possibilities in comparison with the possibilities of UNL. The same framework can be used for building semantic annotations of arbitrary web-documents and web-services and for developing advanced ontologies including the complex definitions of notions. That is why it is possible to say that the purpose of this paper is to bridge a gap between two parallel approaches to adding meaning-understanding ability to the existing web.

The theory of K-representations is a new theory of designing semantic-syntactic analyzers of NL-texts with an extensive use of original formal means for describing the input data and representing the intermediary and final results of semantic-syntactic processing of NL-texts. At the moment, the principal source on this theory is the monograph by Fomichov (2005a). However, many important fragments of this theory are published in English in international editions, in particular, in Fomichov (1996, 1998, 2000, 2002, 2004, 2005b, 2006, 2007). The theory of K-representations is an expansion of the theory of K-calculuses and K-languages (the KCL-theory). The basic ideas and results of the KCL-theory are reflected in numerous publications both in Russian and English, in particular, in Fomichov (1992, 1996, 1998, 2000, 2002, 2004).

The first basic constituent of the theory of K-representations is the theory of restricted standard knowledge languages (RSK-languages), stated, in particular, in Fomichov (1996, 1998, 2000). The kernel of the theory of RSK-languages is a mathematical model introducing a system of such ten partial operations on structured meanings (SMs) of NL-texts that, using primitive conceptual items as "blocks," we are able to build SMs of arbitrary NL-texts (including articles, textbooks, etc.) and arbitrary pieces of knowledge about the world with the help of these ten operations.

The examples considered in this paper use the class of RSK-languages completely defined in Fomichov (1996, 1998). This means that we use all the definitions from Fomichov (1996) with only one exception: we replace the definition of the rule P[9] governing the employment of the universal quantifier and existential quantifier in semantic representations (SRs) of the texts by updated (simplified) definition of the rule P[9] given in Fomichov (1998).

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

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

The second basic constituent of the theory of KRs is a widely applicable mathematical model of a linguistic database (LDB), i.e. a database containing the information used by the algorithms of semantic-syntactic analysis of NL-texts for building SRs of the texts. The model describes the frames expressing the necessary conditions of the existence of semantic relations, in particular, in the word combinations of the following kinds: "Verbal form (verb, gerund) + Preposition + Noun" "Verbal form + Noun" "Noun1 + Preposition + Noun2" "Noun1 + Noun2" "Number designation + Noun" "Attribute + Noun" "Interrogative word + Verb."

The third basic constituent of the theory of KR is a complicated, strongly structured algorithm carrying out semantic-syntactic analysis of texts from some practically interesting sublanguages of NL. This algorithm is based on the elaborated formal model of a LDB. The input texts of this algorithm (the algorithm SemSyn) can be statements, questions of many kinds, and commands. The algorithm SemSyn transforms a NL-text in its SR being a KR, this algorithm is described in two final chapters (12 sections) of the monograph (Fomichov, 2005a). Since, the volume of the description of the algorithm SemSyn is so big, this algorithm cannot be formulated in this paper. It will be important for us here only that this algorithm does exist.

An important feature of the algorithm SemSyn is that it does not construct any syntactic representation of the inputted NL-text but directly finds semantic relations between text units. Since, numerous lexical units have several meanings, the algorithm uses the information from a LDB and linguistic context for choosing one meaning of a lexical unit among several possible meanings.

The other distinguished feature is that a strongly structured algorithm is completely described with the help of formal tools (proposed by the author of this paper), that is why it is problem independent and does not depend on a programming system. The algorithm is implemented in the web programming language PHP.

The structure of this paper is as follows. In Section 2, a conclusion about the restrictedness of expressive possibilities of UNL is drawn. Section 3 explains the mathematical essence of the approach of the theory of RSK-languages to describing SMs of NL texts. In Section 4, the drawbacks of the UNL approach to representing scopes (compound concepts) are analyzed and the advantages of RSK-languages are described. Section 5 sets forth additional principal advantages of RSK-languages in comparison with UNL. The final part of this section contains a proposal to consider the definition of the class of RSK-languages as a model of a SNL of the next generation.Section 6 analyzes the advantages of the theory of KRs in comparison with the principal modern approaches to representing semantic structure of NL-texts: first-order logic, discourse representation theory (DRT), episodic logic (EL), and the theory of conceptual graphs (TCG). In Section 7, the metaphor of a "kitchen combine" is used for summing up a number of new precious opportunities provided by the theory of RSK-languages to the designers of semantic informational technologies. Section 8 discusses the directions of future research work and contains the conclusions. It is proposed to use the definition of the class of RSK-languages not only as a model of a SNL of the next generation but also for:

- building semantic annotations of arbitrary web-documents and web-services; and
- developing advanced intelligent full-text databases and advanced ontologies including, in particular, the complex formal definitions of the notions.

Finally, the conjecture is stated that this paper contributes to bridging a gap between two large-scale parallel approaches to adding the meaning-understanding ability to the existing web: the Semantic Web studies and UNL-based projects.

2. UNL as an initial version of a semantic networking language

UNL represents the NL sentences in the form of formal expressions, without ambiguity, destined not for humans to read, but for computers. The purpose of introducing UNL in communication networks is to achieve accurate exchange of information between different languages. Information has to be readable and understandable by the users. The initial version of UNL (1997-2004) was oriented at representing the contents of only separate sentences but not of discourses. The examples of the basic constructions used in this version of UNL are the expressions:

tower(icl > building), murano(icl > thing-out-of-glass, aoj > colour) build(icl > do), obj(build(icl > do), tower(icl > building)).

where the strings icl, aoj, obj are interpreted as the designations of the binary relations "A concretization of a concept," "An attribute of a thing," "The object of an action." The UNL specifications published in UNL (2005) introduced a manner to build the designations of compound concepts as the so-called scopes (this notion is analyzed in detail below).

The step done in the year 2006 consists in adding the means allowing for representing the meaning of the idioms (UNL, 2006). The analysis shows that in fact the expressive possibilities of UNL are very restricted. First of all, the language UNL is oriented at representing the contents of only separate sentences but not arbitrary discourses. Even the UNL specifications published in 2006 do not contain a theory of representing the meanings of discourses. Besides, UNL is inconvenient for representing, in particular, the meanings of sentences with complicated goals (being parts of advices, commands, wants, etc.), designations of sets, the word "notion," homogeneous members of sentence. Let us consider, for instance, the definition "A flock is a large number of birds or mammals (e.g. sheep or goats), usually gathered together for a definite purpose, such as feeding, migration, or defence." An attempt to represent the meaning of this definition in the language UNL, i.e. with the help of only the designations of binary relations, would lead to a complete destruction of a connection between the structure of the considered definition and the structure of its UNL-representation.

The possibilities of using the language UNL for representing knowledge about the world are very restricted too. Thus, the expressive possibilities of UNL not completely, but only partially correspond to its title "A universal networking language." That is why it seems to be reasonable to interpret the language UNL as not final but only initial version of a SNL. The demands of formally representing the meanings of complicated discourses (for example, pertaining to medicine, science, technology, business, ecology, and low) and the demands of automatic conceptual processing of SRs of such texts with respect to a knowledge base are to lead in the nearest future to the elaboration of a SNL of a new generation. Hence, it is reasonable to look for another, more powerful formal approaches to describing meanings of NL texts with the aim to find (if possible) a model for constructing a universal or widely applicable SNL – understanding ability to the existing web.

3. The essence of a new mathematical approach to describing structured meanings of natural language texts

3.1 An expansion of the universe of formal study

Suppose that we would like to investigate a problem with the help of formal means and, with this aim, to consider a set of entities (real and abstract) and, besides, some relations with the attributes being the elements of these sets and some functions with the arguments and values from this set. Then let us call such a set of entities a universe of formal study.

Imagine, for instance, that we consider the problem of minimizing the cost of delivery of a certain set of goods from a factory to a certain set of shops. Then the universe of formal study includes a factory, the kinds of goods, the concrete goods, the sizes of the goods of each kind, the shops, the lengths of a number of roots, etc.

Under the framework of the first-order predicate logic (FOL), the universe of formal study and the set of formulas describing the properties of the entities from the universe of formal study and the relationships between these entities are two separate sets. It is forbidden, in particular, to construct the formulas of the kind $p(d_1, \ldots d_n)$, where $n \geq 1$, p is an $n-$ary predicate symbol, $d_1, \ldots d_n$ are the attributes of p, and there exists such k, $1 \leq k \leq n$ that d_k is a formula (but not a term). Owing to this restriction, FOL is not convenient, in particular, for expressing the conceptual structure of sentences with direct and indirect speech and with the subordinate clauses of purpose.

The analysis carried out by the author of this paper has shown that a broadly applicable or a universal approach to the formalization of NL semantics is to proceed from a new look at the universe of formal study. We need to expand the universe of formal study by means of adding to the considered set of real and abstract entities (things, situations, numbers, colors, numerical values of various parameters, etc.) the sets consisting of the entities of the following kinds:

- the simple and compound designations of the notions (concepts) qualifying the objects;
- the simple and compound designations of the goals of intelligent systems and of the destinations of things;
- the simple and compound designations of the sets consisting of objects or notions or goals;
- the SRs of the sentences and complicated discourses pertaining to the studied application domains;
- the finite sequences of the elements of any of the mentioned kinds; and
- the mental representations of the NL-texts as informational items having both the content and the metadata (the list of the authors, the date and language of publication, the set of application domains, etc.).

A broadly applicable mathematical framework for the investigation of NL semantics is to allow for considering the relations with the attributes being the elements of an expanded universe of the kind and the functions with the arguments and values from such an expanded universe.

These are just the unique features possessed by the theory of RSK-languages. In particular, this theory allows for including into the universe of formal study the following elements of new kinds:

- A compound designation of a notion:

  book*(Kind1, textbook) (field-of-knowledge, biology);

- A compound designation of a goal of a scholar:

  Defending2*(sci-institution, Lomonosov-Moscow-State-University) (kind-of-dissertation, PhD dissertation*(field-of-knowledge, computer-science)); and

- A compound designation of a set:

  certain art-collection*(quantity, 27) (qualitative-composition, picture).

3.2 The algebraic structure of the basic model describing conceptual operations

During last decade, the most popular approaches to building formal representations of the meanings of NL-texts have been DRT (Kamp and Reyle, 1993, 1996), TCG, represented, in particular, in Sowa (1984, 2000), and EL (Hwang and Schubert, 1993a, b, c; Schubert, 2000; Schubert and Hwang, 2000). In fact, DRT and TCG are oriented at describing the semantic structure of only sentences and short simple discourses. EL studies the structure of only a part of discourses, more exactly, of discourses where the time and causal relationships between the situations (called episodes) are realized.

The analysis shows that the frameworks of DRT, TCG, EL does not allow for considering an expanded universe of formal study satisfying the requirements listed above. That is why the demand to consider an expanded universe of formal study lead the author of this paper in the 1980s and 1990s to the creation of an original mathematical approach to describing conceptual (or semantic) structure of sentences and discourses in NL and operations on conceptual structures needed for building SRs of a broad spectrum of NL-texts.

The definition of the class of RSK-languages (Fomichov, 1996, 1998) became an answer to the following question: how it would be possible to describe in a mathematical way a system of operations on conceptual structures allowing for building (after a finite number of steps) SRs of arbitrarily complicated sentences and discourses from arbitrary application domains, starting from primary informational items.

In other words, an attempt was undertaken to elaborate a new theoretical approach enabling us effectively describe SMs (or contents, or semantic structure, or conceptual structure) of real sentences and arbitrarily complicated discourses pertaining to technology, medicine, business, etc.

As a result, a mathematical model was created describing a system of ten partial operations on SMs of NL-texts and, in particular, determining the class of RSK-languages. The following hypothesis was put forward: using primitive conceptual items as "blocks," we are able to build SMs of arbitrary NL-texts (including articles, textbooks, etc.) and arbitrary pieces of knowledge about the world by means of these ten partial operations. The substantial advantages of RSK-languages in comparison with DRT, TCG, and EL are discussed in Fomichov (1996, 1998, 2002, 2005b).

Let us consider the main ideas of determining a new class of formal languages called RSK-languages. The exact mathematical definitions can be found in Fomichov (1996, 1998). More exactly, we use all the definitions from Fomichov (1996) with only one exception: we replace the definition of the rule P[9] governing the employment of the universal quantifier and existential quantifier in SRs of the texts by updated (simplified) definition of the rule P[9] given in Fomichov (1998).

At the first step (consisting of a rather long sequence of auxiliary steps), a class of formal objects called simplified conceptual bases (s.c.b) is defined. Each s.c.b. B is equivalent to a system of the form $(c_1, \ldots c_{14})$ with the components $c_1, \ldots c_{14}$ being mainly finite or countable sets of symbols and distinguished elements of such sets. In particular, $c_1 = \mathrm{St}$ is a finite set of symbols called sorts and designating the most general considered notions (concepts); $c_2 = P$ is a distinguished sort "sense of proposition"; $c_4 = X$ is a countable set of strings used as elementary blocks for building knowledge modules and SRs of texts; X is called a primary informational universe; $c_5 = V$ is a countable set of variables; $c_7 = F$ is a subset of X whose elements are called functional symbols.

Each s.c.b., B determines three classes of formulas, the first class Lrs(B) being considered as the principal one and being called the RSK-language in the basis B. Its strings (they are called K-strings) are convenient for building SRs of NL-texts. We will consider below only the formulas from the first class Lrs(B).

In order to determine for arbitrary s.c.b. B three classes of formulas, a collection of inference rules P[0], P[1], ... P[10] is defined. The rule P[0] provides an initial stock of formulas from the first class. E.g. there is such a s.c.b. B_1 that, according to P[0], Lrs(B_1) includes the elements:

box1, green, city, set, India, 7, all, any, Weight, Distance, Staff, Suppliers, Quantity, $x1$, \x2, $P5$.

For arbitrary s.c.b. B, let Degr(B) be the union of all Cartesian m-degrees of Lrs(B), where $m \geq 1$. Then the meaning of the rules of constructing well-formed formulas P[0], 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 Lrs(B).

For instance, there is such simplified conceptual basis B that the value of the partial operation Op[7] (it governs the use of logical connectives AND and OR) on the four-tuple:

$< \vee$, Finland, Norway, Sweden $>$

is the K-string:(Finland ∨ Norway ∨ Sweden).Thus, the essence of the basic model of the theory of RSK-languages is as follows: this model determines a partial algebra of the form:

(Degr(B), Operations(B)),

where Degr(B) is the carrier of the partial algebra, Operations(B) is the set consisting of the partial unary operations Op[1], ... Op[10] on Degr(B).

3.3 A short description of the proposed system of conceptual operations

The volume of a concise description in Fomichov (1996) of the mathematical model introducing, in essence, the operations Op[1], ... Op[10] on Degr(B) and, as a consequence, determining the class of RSK-languages is comparable with the volume of this paper. That is why, due to objective reasons, this model cannot be included in this paper. So let us only regard (ignoring many details) the structure of strings which can be obtained by applying any of the rules P[1], ... P[10] at the last step of inferring the formulas.

The rule P[1] enables us to build K-strings of the form Quant Conc where Quant is a semantic item corresponding to the meanings of such words and expressions as "certain," "any," "arbitrary," "each," "all," "several," etc. (such semantic items will be called intensional quantifiers), and Conc is a designation (simple or compound) of a concept. The examples of K-strings for P[1] as the last applied rule are as follows:

certn box1, all box1, certn consignment, certn box1*(Content1, ceramics),

where the last expression is built with the help of both the rules P[0], P[1] and the rule with the number 4, the symbol "certn" is to be interpreted as the informational item corresponding to the word "certain" in cases when this word is associated with singular.

The rule P[2] allows for constructing the strings of the form $f(a_1, \dots a_n)$, where f is a designation of a function, $n \geq 1$, $a_1, \dots a_n$ are K-strings built with the help of any rules from the list P[0], ... P[10]. The examples of K-strings built with the help of P[2]:

Distance(Moscow, Tokyo), Weight(certn box1*(Colour, green)(Content1, ceramics)).

Using the rule P[3], we can build the strings of the form ($a1 \equiv \equiv a2$), where $a1$ and $a2$ are K-strings formed with the help of any rules from P[0], ... P[10], and a_1 and a_2 represent the entities being homogeneous in some sense. The examples of K-strings for P[3] are as follows:

(Distance(Moscow, Tokyo) $\equiv$ $x1$), ($y1 \equiv y3$), (Weight(certn box1) $\equiv$ 8/kg).

The rule P[4] is destined, in particular, for constructing K-strings of the form rel($a_1, \ldots a_n$), where rel is a designation of n-ary relation, $n \geq 1$, $a_1, \ldots a_n$ are the K-strings formed with the aid of some rules from P[0], ... P[10]. The examples of K-strings for P[4]:

Belong(Bonn, Cities(Germany)), Subset(certn series1*(Name-origin, tetracyclin), all antibiotic).

The rule P[5] enables us to construct the K-strings of the form Expr: v, where Expr is a K-string not including v, v is a variable, and some other conditions are satisfied. Using P[5], one can mark by variables in the SR of any NL-text:

- the descriptions of diverse entities mentioned in the text (physical objects, events, concepts, etc.); and
- the SRs of sentences and of larger texts' fragments to which a reference is given in any part of a text.

Examples of K-strings for P[5]: certn box1: $x3$, Higher(certn box1: $x3$, certn box1: $x5$): P1. The rule P[5] provides the possibility to form SRs of texts in such a manner that these SRs reflect the referential structure of NL-texts.

The rule P[6] provides the possibility to build the K-strings of the form ¬ Expr, where Expr is a K-string satisfying a number of conditions. The examples of K-strings for P[6]:

¬ antibiotic, ¬ Belong(penicillin, certn series1*(Name-origin, tetracyclin)).

Using the rule P[7], one can build the K-strings of the forms ($a_1 \wedge a_2 \wedge \ldots \wedge a_n$) or ($a_1 \vee a_2 \vee \ldots \vee a_n$), where $n > 1$, $a_1, \ldots a_n$ are K-strings designating the entities which are homogeneous in some sense. In particular, $a_1, \ldots a_n$ may be SRs of assertions (or propositions), descriptions of physical things, descriptions of sets consisting of things of the same kind, descriptions of concepts. The following strings are examples of K-strings for P[7]:

(streptococcus ∨ staphylococcus), (Belong((Bonn ∧ Hamburg ∧ Stuttgart), Cities(Germany)) ∧ ¬ Belong(Bonn, Cities((Finland ∨ Norway ∨ Sweden)))).

The rule P[8] allows us to build, in particular, K-strings of the form c^*(rel$_1$, val$_1$), ... (rel$_n$, val$_n$), where c is an informational item from the primary universe X designating a concept, for $i = 1, \ldots n$, rel$_i$ is the name of a function with one argument or of a binary relation, val$_i$ designates a possible value of rel$_i$ for objects characterized by the concept c. The following expressions are examples of K-strings for P[8]:

box1*(Content1, ceramics), consignment*(Quantity, 12) (Compos1, box1*(Content1, ceramics)).

The rule P[9] permits to build, in particular, the K-strings of the forms ∀v(conc)D and ∃v(conc)D, where ∀ is the universal quantifier, ∃ is the existential quantifier, conc and D are K-strings, conc is a designation of a primary concept (person, city, integer, etc.) or of a compound concept (integer greater than 200, etc.). D may be interpreted as a SR of an assertion with the variable v about any entity qualified by the concept conc. The examples of K-strings for P[9] are as follows:

∀$n1$(integer)∃$n2$(integer)Less($n1$,$n2$), ∃y(country*(Location, Europe))Greater(Quantity (Cities(y)), 15).

The rule P[10] is destined for constructing, in particular, the K-strings of the form $< a_1, \ldots a_n >$ where $n > 1$, $a_1, \ldots a_n$ are K-strings. The strings obtained with the help of P[10] at the last step of inference are interpreted as designations of n-tuples. The components of such n-tuples may be not only designations of numbers, things, but also SRs of assertions, designations of sets, concepts, etc.

Let us show how it is possible to combine the use of the rules P[0], P[1], ... P[10] while constructing SRs of complex discourses. Let us say that a string Expr of a NL-text T is a KR of T if Expr can be interpreted as a SR of T, and there is such an RSK-language Ls that Expr belongs to Ls.

Example 1. Let T1 = "The chemical action of an electrical current is as follows: for some solutions of acids (or salts, alkalis), while passing a current through such a solution one can observe the isolation of the substances contained in the solution and laying aside these substances on electrodes plunged into this solution. For example, as a result of passing a current across a solution of the blue vitriol ($CuSO_4$), a pure copper will be isolated on the negatively charged electrode. One uses this to obtain pure metals."

Then T1 may have the KR *Expr1* of the form:

(Description(action*(Kind1, chemical)(Agent1, any electr-current), ∃*y*1 (solution1*(Matter1, (acid ∨ salt ∨ alkali))) If-then(Situation(*e*1, passing1*(Agent1, certain electr-current: ?2) (Time, *t*1)(Environment1, *y*1)), Situation(*e*2, observing*(Time, *t*1)(Event1, (certain isolation2*(Agent2, diverse substance1*(Component1, *y*1): *S*1) ∧ certain laying-aside*(Agent2, *S*1)(Loc1, some electrode*(Plunged, *y*1))))))): P1 ∧ Example(*P*1, If-then(Situation(*e*3, passing1*(Agent1, certain electr-current: *y*3) (Time, *t*2) (Environment1, certain solution1*(Matter1, certain substance1*(Isa, blue-vitriol*(Formula1, 'CuSO4')): *y*4))), Situation(*e*4, isolatation2 * (Time, *t*2)(Agent2, certain substance1*(Isa, copper*(Kind, pure)): *y*5)(Loc1, certain electrode*(Charged, negatively): *y*6)))) ∧ Use(certain phenomenon*(Characteristic, *P*1), Our-time, obtaining*(Object1, diverse metal*(Kind, pure): *S*2)))).

Here, the referential structure of T1 is reflected with the help of variables *y*1, *y*2, *y*2, *y*4, *y*5, *y*6, *e*1, *e*2, *e*3, *e*4, *P*1, *S*1, *S*2; the string Expr1 is obtained by applying the inference rules P[0], P[1] (the symbols certain, diverse, some), P[4] (the symbols Description, Kind1, Matter1, If-then, Agent1, Agent2, Isa, Use, etc.), P[5] (the symbol:), P[7] (the logical commectives ∧ and ∨), ?[8] (the symbol *), P[9] (the existential quantifier ∃).

Example 2. The text T1 is taken from the textbook on physics destined for the pupils of the eighth class in Russia (the initial class – six-year-old children – has the number 1, the last class – the number 11). This textbook was written by Pyoryshkin and Rodina and published in Moscow in 1989. This information is reflected by the K-string Expr2 of the form:

certain text*(Content1, Expr2)(Source1, certain text-book*(Educ-institution, any school*(Country, Russia)(Class, 8))(Sci-field, physics)(City, Moscow)(Year, 1989)(Authors, (A.Pyoryshkin ∧ N.Rodina)): inf 517.

where the string inf517 is a mark of a concrete informational object; this mark appeared after the application of the rule P[5] at the final step of constructing this formula.

So we see that RSK-languages allow for building the formulas reflecting both the content of an informational object and its metadata – the data about the informational object as a whole.

4. The advantages of RSK-languages in comparison with UNL concerning the representation of scopes

The analysis shows that it is not difficult to model the basic expressive mechanisms of UNL by means of RSK-languages, because one of the rules used in the definition of the class of RSK-languages is destined for constructing formulas with the names of n-ary relationships and the other rule allows for building compound designations of notions.

Example 1. Let us consider the UNL-expression to(train(icl > thing), London (icl > city)); it denotes a train for London and is taken from Uchida *et al.* (1999). This expression can be approximated by the K-string $E1$ of the form:

Destination(certain train*(Concr, thing), certain city*(Name, 'London')).

or by the K-string $E2$ of the form:

certain train*(Concr, thing)(Destination, certain city*(Name, 'London')).

In a similar way it is possible to construct the K-strings:

tower*(Concr, building), murano*(Concr, thing*(Material, glass)), construction*(Concr, action), Object(certain construction*(Concr, action), certain tower*(concr, building)).

The achieved level of the studies on elaborating the language UNL and using UNL for representing the meanings of regular, non-idiomatic NL-texts is reflected in the monograph (UNL, 2005), having been available online since June 2005. It should be noted that this monograph provides no analysis of the conceptual structure of discourses and, as a consequence, gives no recommendations about the construction of SRs of discourses. The progress achieved in the year 2006 (UNL, 2006) concerns the representation of the meanings of idiomatic expressions. That is why the analysis carried out below is based on the publication (UNL, 2005).

One of the most important steps on the way of expanding the expressive possibilities of the language UNL done in the work (UNL, 2005) in comparison with the previous publications of the authors of this language consists in introducing the notion of a compound concept, or scope.

A scope is defined in UNL (2005, 2006) as a finite set of binary relations grouped for expressing a compound concept. The analysis of the examples illustrating this definition shows that the authors mean distinguishing a set of the expressions of the form $R(c, d)$, where R is the name of a binary relation, c and d are the designations of the attributes of the relation R. The instrument of distinguishing such a set of the formulas is a mark v, inserted into each selected formula immediately after the name of the relation. Thus, a scope is a set of the expressions of the form:

$$R_1 : v(c_1, d_1), \ldots, R_n : v(c_n, d_n), \quad \text{where } n > 1.$$

For example, in UNL (2005, 2006) the sentence Sent1 = (Women who wear big hats in movie theaters) should be asked (to leave) is considered. The part of the sentence within square brackets is interpreted as the expression of what should be asked. The authors construct the following scope reflecting the meaning of two fragments within square brackets:

agt: 01(wear(aoj > thing, obj > hat), woman(icl > person).@pl) obj: 01(wear(aoj > thing, obj > hat), hat(icl > wear)) aoj: 01(big(aoj > thing), hat(icl > wear)) plc:

01(wear(aoj > thing, obj > hat), theater(icl > facilities)) mod: 01(theater(icl > facilities), movie(icl > art)) agt: 01(leave(agt > thing, obj > place).@entry, woman(icl > person).@pl).

The analysis shows that the way of designating compound concepts proposed in UNL (2005, 2006) has the following drawbacks:

- The consideration of two fragments in square brackets as a whole and the construction of the corresponding scope contradict to the language intuition and to the traditions of theoretical and computational linguistics. The sentence Sent1 describes a situation of the type "asking." The following thematic roles (or conceptual cases) are realized in each situation e_k of such type: Agent (a relation between the situation e_k and an intelligent system who is asking); Addressee (a relation between e_k and an intelligent system or a finite set of intelligent systems who are being asked to carry out an action); Goal (a relation between e_k and an action to be performed in order to achieve certain goal). Taking this into account, it is unnatural to build a scope corresponding to the union of the fragments (Women who wear big hats in movie theaters) and (to leave). Both our language intuition and the traditions of theoretical and computational linguistics show the expedience of constructing a scope for the fragment (Women who wear big hats in movie theaters).
- An essential drawback of the notion of a scope is that the form of a scope does not allow for distinguishing a scope designating certain set of entities from a scope designating the content of a complicated phrase or a discourse. Meanwhile, the discourses very often contain the expressions referring to the meaning of preceding or following phrases or larger parts of the text. That is why it is necessary to elaborate the means of expressing such references in SRs of discourses.
- The next drawback of the notion of a scope is as follows. There are possible the situations when the set of the attributes of the relations belonging to a scope contains two or more entities qualified by the same concept (e.g. three ships). However, the monographs (UNL, 2005, 2006) do not contain any analysis of such situations and give no recommendations about how to designate various entities qualified by the same concept and belonging to one scope.

The theory of KRs provides such means of constructing the designations of compound concepts which allow for escaping the mentioned drawbacks of the notion of a scope in the specifications of the language UNL.

Example 2. In order to describe the set of all women wearing big hats at a certain moment in a certain cinema, it is possible to use the following K-string:

certain set*(Quality-composition, woman)(Set-mark, *S*1)(set-description, properties-of-elements (arbitrary woman*(element, *S*1):*y*1, situation(*e*1, dress-wearing*(agent1, *y*1)(object1, certain hat*(size, big):*x*1)(place, certain cinema:*x*2)(time, *t*1)))).

5. Additional principal advantages of RSK-languages in comparison with UNL

The theory of RSK-languages possesses many precious properties being additional important advantages concerning the construction of a SNL of a new generation in comparison with UNL. Let us illustrate a number of such properties.

Property 1. The possibility to build compound designations of notions (concepts).

Example 1. Let T1 = "A flock is a large number of birds or mammals (e.g. sheep or goats), usually gathered together for a definite purpose, such as feeding, migration, or defence." T1 may have the KR Expr1 of the form:

Definition1(flock, dynamic-group*(qualitative-composition, (bird ∨ mammal*(Examples, (sheep ∧ goal)))), *S*1, (Estimation1(Quantity(*S*1), high) ∧ Goal-of-forming(*S*1, certain purpose*(Examples, (feeding ∨ migration ∨ defence))))).

The analysis of this formula enables us to conclude that it is convenient to use for constructing SRs of NL-texts:

- the designation of a five-ary relationship Definition1;
- compound designations of concepts (in this example the expressions mammal* (examples, (sheep ∧ goal)) and dynamic-group*(qualitative-composition, (bird ∨ mammal*(examples, (sheep ∧ goal)))) were used);
- the names of functions with the arguments and/or values being sets (in the example, the name of an unary function Quantity was used, its value is the quantity of elements in the set being an argument of this function); and
- compound designations of intentions, goals (in this example it is the expression:

certain purpose*(Examples, (feeding ∨ migration ∨ defence)).

The structure of the constructed KR Expr1 to a considerable extent reflects the structure of the definition T1. Meanwhile, any attempt to represent the content of this definition in the language UNL, i.e. with the help of only binary relationship, would destroy any similarity between the structure of T1 and the structure of its UNL-representation.

Property 2. The possibility to build compound designations of goals.

Example 2. Let T2 = "The policyholder phones Europe Assist to inform about a car damage". Then T2 may have the following KR (KR), i.e. a SR being an expression of some RSK-language:

Situation(*e*1, phone-communic*(Agent1, certn person*(Hold1, certn policy1:*x*1):*x*2)(Object2, certn firm1*(Name1, "Europe Assist"):*x*3)(Purpose, inform-transfer*(Theme1, certn damage1*(Object1, certn car1):*x*4))).

Property 3. The possibility to construct compact SRs of such fragments of sentences which are obtained by means of joining the designations of things, events, concepts or goals with the help of logical connectives ∧ (AND), ∨ (OR).

Example 3. Let T3 = "After receiving a repair invoice from the firm 'Lee C.S.' and a claim from the policyholder, the company 'AGFIL' pays the car repair to the garage." Then a KR of T3 can be the expression:

(Situation(*e*1, (receiving1*(Agent2, certn firm1*(Name1, "AFGIL"): *x*1) (Object1, certn invoice*(Theme, certn repair: *e*2): *x*2)(Sender1, certn firm1*(Name1, "Lee C.S."): *x*3) ∧ receiving1*(Agent2, *x*1) (Object1, certn claim1: *x*4) (Sender1, certn person*(Hold1, certn policy1: *x*5): *x*6))) ∧ Situation(*e*2, payment1*(Agent2, *x*1)(Addressee1, certn garage: *x*7)(Sum, Cost(*e*2))) ∧ Before(*e*1, *e*2)).

Property 4. The existence of the formal means allowing for representing SMs of the discourses with the references to the meanings of sentences and larger fragments of the texts.

Example 4. Let T4 = "All granulocytes are polymorphonuclear; that is, they have multilobed nuclei." Then T4 may have the following KR Expr:

(Property(arbitrary granulocyte:$x1$, polymorphonuclear): P1 ∧ Explanation(P1, If-then(Situation($e1$, possessing1*(Subject1, $x1$)(Object1, certain nucleus:$x2$)), Property($x2$, multilobed)))).

Here, $P1$ is the variable marking the meaning of the first phrase of T4; the strings Subject1, Object1 designate thematic roles (or conceptual cases).The key role in the construction of the KR Expr was played by the rule enabling to introduce the mark $x1$ for designating an arbitrary granulocyte, the mark $x2$ for designating the nucleus of the cell $x1$, and the mark $P1$ for designating SR of the first sentence from the discourse $T2$. The mark (variable) $P1$ enables to explicate in the structure of SR of T4 the reference to the meaning of the first sentence; this reference is given by the word combination "that is".

The language UNL does not provide the means for representing the meanings of sentences and larger fragments of discourses. Meanwhile, the last example considers one of the shortest discourses of the kind. The textbooks in various fields of knowledge contain a lot of much more complicated discourses with the references to the meanings of sentences and larger fragments of discourses.

Example 5. Let T5 = "The firm "Europe Assist" provides a policyholder with a telephone service; in particular, assigns a garage for repair and informs the company "AGFIL" about a claim of a policyholder.. Then T5 may have a KR:

(Situation($e1$, service1*(Agent2, certn firm1*(Name1, "Europ Assist"): $x1$)(Instrument, certn telephone: $x2$)(Object1, arbitrary person**(Hold1, certn policy1: $x3$): $x4$): $P1$ ∧ Concretization($P1$, ((Situation($e2$, assigning1*(Agent2, $x1$)(Addressee1, $x4$))(Object3, certn garage*(Destination1, repair): $x5$)) ∧ Situation($e3$, inform-transfer*(Agent2, $x1$)(Addressee1, certn firm1*(Name1, "AFGIL"): $x6$)(Content1, certn claim1*(Authors, $x4$): $x7$))))).

The variable $P1$ in the constructed formula is a mark of the SR of the sentence S1 = "The firm 'Europ Assist' provides a policyholder with a telephone service." In the second part of the discourse T5, this mark is used for representing in a compact way the reference to the meaning of the sentence $S1$.

Property 5. The existence of formal means allowing for constructing compound designations of sets as components of SRs of NL-texts.

Example 6. A set consisting of 12 single rooms in the three-star hotels of Vienna may have a KR of the form:

certn set*(Number, 12)(Qualitative-composition, room*(Kind1, single)(Location, any hotel*(Kind2, three-star)(Loc, Vienna))).

Property 6. The possibility to build object-oriented SRs of informational objects, in particular, of the records of negotiations or contracts, i.e. the expressions of the form:

certn inform-object*(Kind, concept)(Content1, cont)$(r_1, u_1)\ldots(r_n, u_n)$.

For instance, if we consider the records of negotiations or contracts, concept is the designation of the notion "negotiation record" or "contract" cont is a KR of a document, $r_1, \ldots r_n$ are the designations of the external characteristics of a document (expressing its metadata, for instance, the data about the authors, date, language, etc.), and $u_1, \ldots u_n$ are the strings interpreted as the designations of the data associated with a document.

The considered examples show that RSK-languages enable us, in particular, to describe the conceptual structure of texts with:

- references to the meanings of phrases and larger parts of texts;
- compound designations of sets;
- definitions of terms;
- complicated designations of objects; and
- generalized quantifiers (arbitrary, certain, etc.).

Besides, RSK-languages provide the possibilities to describe the semantic structure of definitions, to build formal analogues of complicated concepts, to mark by variables the designations of objects and sets of objects, to reflect thematic roles.

The additional useful properties of RSK-languages from the standpoint of building SRs of NL-texts are the possibilities:

- to explicitly indicate thematic roles (or conceptual cases, or semantic cases) in the structure of SRs of NL-texts;
- to reflect the meanings of the phrases with direct and indirect speech, with the word "a concept"; and
- to consider the functions with the arguments and/or values being the sets of objects or concepts (Suppliers, Staff, etc.).

The creation of a SNL belonging to a new generation on the basis of the definition of the class of RSK-languages, in particular, will allow for:

- constructing not only SRs of separate sentences but also SRs of complicated discourses with the help of reflecting the references to the previously mentions entities and to the meanings of phrases and larger fragments of discourses;
- forming compound designations of sets, concepts, goals of intelligent systems and destinations of things;
- joining with the help of logical connectives "and" "or" not only the designations of assertions (as in predicate logic) but also the designations of concepts, objects, sets of objects;
- reflecting the semantic structure of the phrases with the words "concept" "notion"; and
- considering non-traditional functions with arguments and/or values being sets of objects, sets of concepts, SRs of texts, sets of SRs of texts.

Thus, the theory of KRs opens the real prospects of constructing a SNL of a new generation with the expressive possibilities being much closer to the expressive possibilities of NL in comparison with the language UNL (2005, 2006).

In Fomichov (1996, 2005b), the hypothesis is formulated that the theory of RSK-languages provides the effective means for describing SMs (i.e. for representing contents) of arbitrary NL-texts in arbitrary thematic domains. That is why the following conjecture seems to be well grounded: the theory of KRs provides a model of a universal SNL.

6. Related approaches to representing semantic structure of NL-texts

During the last decade, the most popular approaches to representing semantic structure of NL-texts were first-order logic, DRT (Kamp and Reyle, 1993, 1996), EL (Hwang, 1992; Hwang and Schubert, 1993a, b, c; Schubert, 2000; Schubert and Hwang, 2000), and the TCG (Sowa, 1984, 2000).

The advantages of the theory of RSK-languages in comparison with first-order logic, DRT and EL are, in particular, 11 possibilities to:

(1) distinguish in a formal way objects (physical things, events, etc.) and concepts qualifying these objects;

(2) build compound representations of concepts;

(3) distinguish in a formal manner objects and sets of objects, concepts and sets of concepts;

(4) build complicated representations of sets, sets of sets, etc.;

(5) describe set-theoretical relationships;

(6) effectively describe SMs of discourses with references to the meanings of phrases and larger parts of discourses;

(7) describe SMs of sentences with the words "concept," "notion";

(8) describe SMs of sentences where the logical connective "and" or "or" joins not the expressions-assertions but designations of things, sets, or concepts;

(9) build complicated designations of objects and sets;

(10) consider non-traditional functions with arguments or/and values being sets of objects, of concepts, of texts' SRs, etc.; and

(11) construct formal analogues of the meanings of infinitives with dependent words and, as a consequence, to represent proposals, goals, obligations, commitments.

It should be added that the model has at least three global distinctive features as concerns its structure and destination in comparison with EL. The first feature is as follows. In fact, the purpose of the paper (Fomichov, 1996) is to represent in a mathematical form a hypothesis about the general mental mechanisms (or operations) underlying the formation of complicated conceptual structures (or semantic structures, or knowledge structures) out of basic conceptual items. EL does not undertake an attempt of the kind, and 21 Backus-Naur forms used in Hwang (1992) for defining the basic logical syntax rather disguise such mechanisms (operations) in comparison with more general ten rules described in Fomichov (1996).

The second global distinctive feature is that the paper (Fomichov, 1996) formulates a hypothesis about a complete collection of operations of conceptual level providing the possibility to build effectively the conceptual structures corresponding to arbitrarily complicated real sentences and discourses pertaining to science, technology, business, medicine, law, etc.

The third global distinction is that the form of describing in EL the basic collection of informational items is not a strictly mathematical one. E.g. the collection of Backus-Naur forms used with this purpose in Hwang (1992) contains the expressions:

< *l*-place-pred-const > :: = happy|person|certain|probable| ...
<1-fold-pred-modifier-const> :: = plur|very|former|almost|in-manner|...

The only way to escape the use of three dots in productions is to define some analogue of the notion of a simplified conceptual basis introduced in the paper (Fomichov, 1996).

The items (3)-(8), (10), (11) in the list above indicate the principal advantages of the theory of SK-languages in comparison with the TCG. Besides, the expressive possibilities of the new theory are much higher than the possibilities of TCG as concerns the items (1), (2), (9). In particular, TCG does not allow for building an analogue of the constructed above semantic description of a set consisting of 12 single rooms in the three-star hotels of Vienna.

The author of this paper carried out a comparative analysis of the expressive possibilities of SK-languages and of the NL phenomena reflected in the structure of commercial contracts and the records of negotiations. A number of the results of this analysis is set forth in Fomichov (2005b, 2006). The fulfilled analysis allows for formulating the assumption that the expressive possibilities of RSK-languages are sufficient for building with their help the formal representations of contracts and records of commercial negotiations.

On the other hand, the expressive possibilities of another known formal approaches to representing the contents of NL-texts are insufficient for building SRs of arbitrary contracts and records of negotiations. In particular, it applies to first-order logic, DRT, TCG, and EL.

7. A kitchen combine for the designers of semantic informational technologies as a metaphor concerning the theory of RSK-languages

It seems that a metaphor can help to better grasp the significance of the theory of RSK-languages for the designers of semantic informational technologies, first of all, for the designers of NLPSs. It establishes a connection between the problems of a house-keeping and the problems associated with the development of semantic informational technologies.

When a woman having a full-time job enters the kitchen, she has a lot of things to do in a short time. The kitchen combines are constructed in order to make the work in the kitchen easier and to diminish the time needed for the work of the kind. For this, the kitchen combines can chop, slice, stir, grate, blend, squeeze, grind and beat.

The designers of NLPSs have a lot of things to do in a very restricted time. That is why they need effective formal tools for this work. Like a kitchen combine for house-keeping, the theory of RSK-languages can help the designers of semantic informational technologies to do many things. In particular, the theory of RSK-languages is convenient for:

- constructing formal definitions of concepts;
- representing knowledge associated with concepts;

- building knowledge modules (in particular, definitions of concepts) as the units having both the content (e.g. a definition of a concept) and the external characteristics (e.g. the authors, the date of publishing, the application fields);
- representing the goals of intelligent systems;
- building SRs of various algorithms given in a NL form;
- representing the intermediate results of semantic-syntactic processing of NL-texts (in other words, building underspecified SRs of the texts);
- forming final SRs of NL-texts;
- representing the conceptual macro-structure of a NL discourse;
- representing the speech acts; and
- building high-level conceptual descriptions of the figures occurred in the scientific papers, textbooks, technical patents, etc.

No other theory in the field of formal semantics of NL can be considered as a useful tool for all enumerated tasks. In particular, it applies to Montague Grammar and its extensions and to DRT, TCG and EL.

In case of technical systems, a highly precious feature is the simplicity of construction. Very often, this feature contributes to the reliability of the system and the easiness of its exploitation. The theory of RSK-languages satisfies this criterion too, because it makes the following discovery both in non-mathematical and mathematical linguistics: a system of such 10 operations on SMs of NL-texts is found that, using primitive conceptual items as "blocks," we are able to build SMs of arbitrary NL-texts (including articles, textbooks, etc.) and arbitrary pieces of knowledge about the world. Such operations will be called quasilinguistic conceptual operations. Hence, the theory of KRs suggests a complete collection of quasilinguistic conceptual operations (it is a hypothesis supported by many weighty arguments).

The useful properties of RSK-languages stated above allow for the conclusion that the theory of RSK-languages can be at least not less useful for the designers of NLPSs and for a number of other semantic informational technologies as a kitchen combine is of use for making easier the work in the kitchen.

It may be added that the works (Fomichov, 2005a, b, 2006) ground the possibility of using the theory of RSK-languages for reflecting the contents of the records of commercial negotiations conducted by computer intelligent agents (CIAs) and for building the contracts concluded by CIAs. In fact, the elaboration of the definition of the class of RSK-languages means that a class of formal languages is determined being convenient for building semantic descriptions of arbitrary goods, services, and contracts. It seems that now the theory of SK-languages is the only formal theory allowing for constructing semantic analogues of arbitrary complicated goals and, as a consequence, of offers, promises, commitments. That is why its potential of developing general-purpose formal languages of business communication exceeds the potential of first-order logic, DRT, TCG, and EL.

It is stressed in Hasselbring and Weigand (2001) that it is particularly interesting for e-commerce that one format can be used both for electronic messages (to be processed by computers) and for human interfaces. Many years experience of using RSK-languages for the design of NL-interfaces show that RSK-languages possess this precious feature.

One of other possible applications of the new formal theory in e-commerce can be the elaboration of feedback languages for knowledge acquisition by CIAs, i.e. the use of RSK-languages for representing in a formal (but readable for the end-users) way the meanings of NL descriptions of knowledge fragments inputted into computer by the end-user.

8. Conclusion and future work

The analysis of expressive power of the language UNL provides the possibility to establish an analogy between the studies on constructing a SNL (UNL being one of its versions) and the researches on the development of the advanced informational languages for forming web-documents. The conclusion can be drawn that, similar to the ongoing process of the transition from the language HTML to new, semantically-structured means for representing information on the web, in the field of constructing a SNL the demands of practice must lead in the coming years to the creation of a SNL belonging to a new generation in comparison with UNL.

The prospects of using the theory of KRs for the elaboration of a SNL with the expressive power exceeding the expressive power of UNL are set forth. The hypothesis is put forward that the definition of the class of RSK-languages can be used as a model for the development of a universal SNL.

The examples considered above show that the UNL-expressions describing various entities and binary relations between these entities may be easily transformed into the K-strings (i.e. the expressions of SK-languages). That is why the linguistic processors designed in many countries under the framework of the UNL project can be easily modified in order to build K-strings. Thus, the choice of the theory of KRs for continuing the studies in this direction is practically possible.

The principal advantages of such a choice are as follows:

- In comparison with the publications on UNL, the theory of KRs provides a powerful and flexible framework for designing semantic-syntactic analyzers of arbitrarily complicated discourses but not only of separate sentences.
- The theory of KRs gives the formal means being convenient for building SRs of definitions of notions and sentences with the descriptions of sets, compound descriptions of goals and compound destinations of things.

Concerning the directions of future research work, it seems to be worthwhile to use the definition of the class RSK-languages for building semantic annotations of arbitrary web-documents and of arbitrary web-services and to start the studies aimed at the elaboration of the effective methods of intelligent processing of such semantic annotations.

The algorithm of semantic-syntactic analysis of NL-texts SemSyn, completely described in Fomichov (2005a), can become a good starting point for developing the algorithms and computer programs transforming the NL annotations of web documents into their semantic annotations.

Besides, the theory of RSK-languages can be used for developing intelligent full-text databases and advanced ontologies, including the complex definitions of notions. That is why it seems to be reasonable to believe that this paper contributes to bridging a gap between two parallel approaches to adding meaning-understanding ability to the existing web.

References

Fomichov, V.A. (1992), "Mathematical models of natural-language-processing systems as cybernetic models of a new kind", *Cybernetica (Belgium)*, Vol. 35 No. 1, pp. 63-91.

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, pp. 5-32.

Fomichov, V.A. (1998), "Theory of restricted K-calculuses as a comprehensive framework for constructing agent communication languages", *Informatica. An International Journal of Computing and Informatics (Slovenia, Europe)*, Vol. 22, Special Issue on NLP and Multi-Agent Systems, ed. by V.A. Fomichov and A.P. Zeleznikar, pp. 451-63.

Fomichov, V.A. (2000), "An ontological mathematical framework for electronic commerce and semantically-structured web", in Zhang, Y., Fomichov, V.A. and Zeleznikar, A.P. (Eds), *Informatica. An International Journal of Computing and Informatics (Slovenia, Europe)*, Vol. 24, Special Issue on Database, Web, and Cooperative Systems, pp. 39-49.

Fomichov, V.A. (2002), "Theory of K-calculuses as a powerful and flexible mathematical framework for building ontologies and designing natural language processing systems", in Andreasen, T., Motro, A., Christiansen, H. and Larsen, H.L. (Eds), *Flexible Query Answering Systems, 5th International Conference, FQAS 2002, Copenhagen, Denmark, October 27-29, 2002. Proceedings; LNAI 2522 (Lecture Notes in Artificial Intelligence, Vol. 2522)*, Springer, New York, NY, pp. 183-96.

Fomichov, V.A. (2004), "Theory of standard K-languages as a model of a universal semantic networking language", *Proceedings "Intelligent Software Systems for the New Infostructure" (Focus Symposium in Conjunction with the 16th Intern. Conf. on Systems Research, Informatics and Cybernetics – InterSymp-2004, July 29 – August 5, 2004, Germany). Focus Symposia Chair: Jens Pohl.*, CAD Research Center, Cal Poly, San Luis Obispo, CA, pp. 51-61.

Fomichov, V.A. (2005a), *The Formalization of Designing Natural Language Processing Systems*, Max Press, Moscow.

Fomichov, V.A. (2005b), "Standard K-languages as a powerful and flexible tool for building contracts and representing contents of arbitrary E-negotiations", in Bauknecht, K., Proell, B. and Werthner, H. (Eds), *The 6th Intern. Conf. on Electronic Commerce and Web Technologies "EC-Web 2005", Copenhagen, Denmark, August 23-26, 2005, Lecture Notes in Computer Science*, Vol. 3590, Springer-Verlag, New York, NY, pp. 138-47.

Fomichov, V.A. (2006), "The theory of SK-Languages as a powerful framework for constructing formal languages for electronic business communication", *Proceedings "Advances in Intelligent Software Systems" (Focus Symposium in Conjunction with the 18th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2006, August 7-12, 2006, Baden-Baden, Germany), Focus Symposia Chair: Jens Pohl.*, CAD Research Center, Cal Poly, San Luis Obispo, CA.

Fomichov, V.A. (2007), "Theory of K-representations as a framework for constructing a semantic networking language of a new generation", *Proceedings "Representation of Context in Software: Data = > Information = > Knowledge" (Focus Symposium in Conjunction with the 19th International Conference on Systems Research, Informatics and Cybernetics – InterSymp-2007, July 30-August 4, 2007, Germany). Focus Symposia Chair: Jens Pohl.*, CAD Research Center, Cal Poly, San Luis Obispo, CA, pp. 61-70.

Hasselbring, W. and Weigand, H. (2001), "Languages for electronic business communication: state of the art", *Industrial Management & Data Systems*, Vol. 101, pp. 217-26.

Hwang, C.H. (1992), "A logical approach to narrative understanding", PhD dissertation, University of Alberta, Edmonton.

Hwang, C.H. and Schubert, L.K. (1993a), "Meeting the interlocking needs of LF-computation, deindexing, and inference: an organic approach to general NLU", *Proc. 13th Int. Joint Conf. on AI (IJCAI'93), Chambery, France-1993, 28 August-3 September*.

Hwang, C.H. and Schubert, L.K. (1993b), "Episodic Logic: a situational logic for natural language processing", in Aczel, P., Israel, D., Katagiri, Y. and Peters, S. (Eds), *Situation Theory and its Applications/CSLI Lecture Notes*, Vol. 3, CSLI Publications, Stanford, MA, pp. 303-38.

Hwang, C.H. and Schubert, L.K. (1993c), "Episodic logic: a comprehensive, natural representation for language understanding", *Minds and Machines*, Vol. 3, pp. 381-419.

Kamp, H. and Reyle, U. (1993), *Introduction to Modeltheoretic Semantics of Natural Language, Formal Logic and Discourse Representation Theory*, Kluwer Academic Publishers, Dordrecht.

Kamp, H. and Reyle, U. (1996), "A calculus for first order discourse representation structures", *Journal for Logic, Language and Information*, Vol. 5, pp. 297-348.

Schubert, L.K. (2000), "The situations we talk about", in Minker, J. (Ed.), *Logic-Based Artificial Intelligence*, Kluwer Academic Publishers, Dortrecht, pp. 407-39.

Schubert, L.K. and Hwang, C.H. (2000), "Episodic logic meets little red riding hood: a comprehensive, natural representation for language understanding", in Iwanska, L. and Shapiro, S.C. (Eds), *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language*, MIT Press, Menlo Park, CA, pp. 111-74.

Sowa, J.F. (1984), *Conceptual Structures: Information Processing in Mind and Machine*, Addison-Wesley, Reading, MA.

Sowa, J.F. (2000), *Knowledge Representation: Logical, Philosophical, and Computational Foundations*, Brooks/Cole, Pacific Grove, CA.

Uchida, H., Zhu, M. and Della Senta, T. (1999), *A Gift for a Millennium*, A book published by the United Nations University, Tokyo; available at: www.unl.ias.unu.edu/publications/gm/index.htm

UNL (2005), Universal Networking Language (UNL) Specifications, Version 2005, available at: www.undl.org/unlsys/unl/unl2005/

UNL (2006), *Universal Networking Language (UNL) Specifications, Version 2005, Edition 2006, 30 August 2006*, UNL Center of UNDL Foundation, available at: www.undl.org/unlsys/unl/unl2005-e2006/

About the author

Vladimir Fomichov is a Professor with the Department of Innovations and Business in the Sphere of Informational Technologies, Faculty of Business Informatics, State University-Higher School of Economics, Moscow, Russia. He is also a Half-time Professor of Computer Science at the Department of Information Technologies, K.E. Tsiolkovsky Russian State Technological University, Moscow. He is a member of the International Society for Artificial Intelligence in Education (Head Office in Leeds, UK) since 1996, and also the Editorial Board Member of *Informatica*, since 1992. He has authored and co-authored over 125 scientific publications in Russian, English, French (including three monographs) pertaining to the fields of discrete mathematics, artificial intelligence, multi-agent systems, semantic web, electronic commerce, and cognitive science. Vladimir Fomichov can be contacted at: vdrfom@aha.ru. For further information, please visit the web site: http://new.hse.ru/C3/C0/Fomichev/default.aspx