Bull. Nov. Comp. Center, Comp. Science, 26 (2007), 147–158

Logical analysis of texts in a natural language and a sense representation

Tatyana Batura, Feodor Murzin

Abstract. Methods for analysis of texts and separate sentences in a natural language are under discussion. Their main application is to study a written speech with the help of mathematical logic, syntactic rules and the morphology of the modern Russian literary language. Various algorithms for matching predicates and formulas of first order predicate calculus with natural language texts are considered. Also some finite models are matched with parts of text and the text as a whole. The results of this paper may be applied in computer-aided systems of extracting information from natural language texts, in intellectual systems of searching information in the Internet, etc.

1. Introduction

Within frameworks of the realized project, it is supposed to develop methods which will allow to carry out miscellaneous analysis of texts and separate sentences in a natural language. Such methods are planned to use as the sense representation for a text in a natural language within frameworks of Melchuk's approach and lexical functions proposed by him [1, 2], Markus's set-theoretical models [3], and to adapt some methods and constructions of mathematical logic for analyzing texts in a natural language, in particular Henkin's construction used for proving the model existence theorem and the omitting types theorem [4], finite forcing etc.

The purpose of this work is developing various algorithms for matching predicates and formulas of first order predicate calculus with natural language texts. The authors have also considered a possibility to match finite models with text sentences and even the whole text.

In the future the achieved results may be studied and transformed by mathematical logic means. It gives us an opportunity to make a transition from a syntactic to a semantic level and in some sense teach a machine to understand a meaning of a natural language text.

The results of this work may be applied in computer-aided systems of extracting information from natural language texts, in intellectual systems of searching information in the Internet, in constructing automated summarizing systems, electronic translators and dictionaries.

The given work may also be useful for developing various search systems, in cases when it is needed to extract necessary information from a document

by query or to select required documents from large amount of documents by the given query. On basis of this work it will be possible to develop systems that will be able to reconstruct text sense and to extract knowledge from the text that may be presented to a user in the form of compact reports (schemes, abstracts) or referred to the knowledge base.

2. A review of methods for representing natural language text meaning

One of the matching predicates algorithms is based on lexical functions proposed by Melchuk. On syntactic level, these functions may be represented as predicates in the following form. One can consider the whole set of word forms in a language which appear when nouns are declined, verbs are conjugated etc. (i.e. the whole vocabulary) and suppose that x and y are words or word combinations from this set. Then we have predicates of the following form:

$Syn(x, y)$, x, y are synonyms;
$Anti(x, y)$, x, y are antonyms;
$Conv(x, y)$, x, y are conversives;
$Gener(x, y)$, y is a standard name for a concept, patrimonial in relation to a concept designated by x ($x =$ "strawberry", $y =$ "berry");
$Destr(x, y)$, y is a standard name for an "aggressive" action ($x =$ "bee", $y =$ "stings").
$Doc(x, y)$, y is a "document";
$Doc_{res}(x, y)$, y is a document, as a result, ($x =$ "to report", $y =$ "the report");
$Doc_{perm}(x, y)$, y is a document on the right, ($x =$ "train", $y =$ "the travel card");
$Doc_{cert}(x, y)$, y is a document certifying, ($x =$ "higher education", $y =$ "diplom").

We can right down the formula
$(\forall x)(\forall y)\,(Doc(x, y) \leftrightarrow Doc_{res}(x, y) \vee Doc_{perm}(x, y) \vee Doc_{cert}(x, y))$.

The Markus set-theoretical models are constructed as follows. He considers some class decomposition of a natural language vocabulary (which is supposed to be a finite set). For example this decomposition may consist of classes corresponding to inflectional word-form sets. With the help of such decomposition it is possible to give a formal definition of gender and case. Markus also defines the so-called "syntactic types" which correspond practically with the traditional parts of speech. On the basis of syntactic types operations there is a possibility to establish grammatical correctness of a natural language sentence.

Representation of semantic structure of texts in a natural language by means of formulas of first order logic was considered also in articles

of Paducheva [5], Lakhuti and Rubashkin [6]. There are also the similar works fulfilled in Institute of Mathematics SBRAS and in Novosibirsk State University. For example, articles of Palchunov, Moskvitin, Zagoruiko, Churikova, Vikentiev.

Our concept consists of using Henkin's construction from a mathematical logic to construct some finite models, which can be treated as a sense of text.

3. Structures corresponding to natural language sentences

A part of the carried out work may be described as follows. Several structures $structure_1, ..., structure_q$ will correspond to each sentence and a set of predicates $predicate_{i1}, ..., predicate_{ij(i)}$ corresponds to each structure $structure_i$.

On the other hand it is also possible to consider elements of natural language vocabulary as constants, then introduce predicates and get formulas on their basis. Further at first the predicates can be considered on a syntactic level. After that they are considered as subsets of basic model sets in corresponding Cartesian powers. This approach gives a possibility to construct models, i.e. to realize a transition from syntactic to semantic level.

As an example, let's consider structures that correspond with verbs. They may be obtained in the following way. Let us suppose that there is only one verb and several nouns in different cases (which are related to this verb) in the sentence. Every case is considered to have no more than one noun. Such sentence may be matched with the following structure

V	NNom	NGen	NDat	NAcc	NInstr	NPrep

where *NPrep* is a noun in prepositional case (if there is any), etc. When there is no noun in this case in the sentence the corresponding position of the structure may be filled up by some auxiliary information about the fact that there is no noun in this case in the sentence but in principal it can be placed there, or there is no noun in the given case and it cannot exist there at all.

The predicate $P(v, n_1, ..., n_6)$ corresponds to this structure where v is a verb, $n_1, ..., n_6$ are nouns. The predicate is sextiary since there are six cases in Russian.

Another example. Let us assume that there are some prepositions in the sentence. First of all we have to define for every preposition to which noun it relates.

The simplest case when a preposition is placed immediately near a noun or it is separated from the noun with the help of one or several adjectives. Then we just add some prepositions to the structure.

If as a result prepositions and nouns are correlated to each other, the following predicate appears: $P(v, prep_{11}, n_{11}, ..., prep_{1k}, n_{1k}; ...; prep_{61},$

$n_{61}, ..., prep_{6k}, n_{6k}$) where $prep_{ij}$ are prepositions. If there is no any preposition in the sentence then we fix this fact with the help of a constant, for instance *Nil*. In particular there is no any preposition related to a noun in nominative case.

More complicated situation is appearing if a preposition is separated from a noun by an "extended attribute". In this case the question about agreement of a preposition and a noun in any case is considered. Nevertheless if one cannot establish any connection of a preposition with a noun then it is possible to resort to the frequency compatibility of words.

4. Grammatical predicates

There is one more way of introducing predicates — matching with parts of speech. We call such predicates as **grammatical predicates**. For example $N(x, y_1, ..., y_n)$, x is a noun, y_i are characteristics used for dividing nouns into several groups. The notation $N(x, y_1, ..., \underset{i}{0}, ..., y_n)$ means an absence of i-th characteristic.

Let the subscript of a predicate corresponds to an order number of the indicated characteristic and the superscript corresponds to the number of the property rendering this characteristic.

If characteristics $y_1, ..., y_n$ are alternative, we will denote this by $N(x, y)$, where $y = y_1$, if x has a characteristic y_1;...; $y = y_n$, if x has a characteristic y_n.

Let us consider the following example, namely a noun number, singular or plural forms. It is an alternative characteristic, since nouns cannot be in singular and plural form at the same time. However the noun can exist in different cases simultaneously (*метро*), have masculine and feminine gender (*плакса*), be animate and inanimate (*пень*) etc. We do not regard these characteristics as alternative ones.

Because of this, the XOR operation is defined differently. For $P(x, y_1, ..., y_n)$-type predicates the XOR is defined as conjunction of disjunctions, for example:

$Prep_1(x, y)$ means that prepositions are divided by their origin into $y =$ "непр", i.e. x is an underivative (prototypal) preposition and $y =$ "пр", i.e. x is a derivative preposition. Derivative prepositions are divided into

a) $Prep_1^1(x)$ — derived from an adverb (adverbial) (*близ, около, сквозь* etc.);

b) $Prep_1^2(x)$ — derived from a noun (nounal) (*вследствие, по пути, по причине* etc.);

c) $Prep_1^3(x)$ — derived from a verb (verbal) (*благодаря, спустя* etc.).

From here we obtain

$$(\forall x)\left(Prep_1(x, np) \leftrightarrow \mathop{\&}_{i=1,\ i\neq j}^{3}\left(\left(Prep_1^i(x)\&\neg Prep_1^j(x)\right) \vee \left(Prep_1^j(x)\&\neg Prep_1^i(x)\right)\right)\right).$$

This formula means that a preposition is the derivative preposition if and only if it contains to one of the groups a)–c).

For the predicates of the type $P(x, y)$ this operation matches with the usual "or". For example:
$N_5(x, y)$, y = "отвл", if the noun is abstract, y = "конкр", if the noun is concrete (they represent individual objects, living creatures and some phenomena of environment).
$(\forall x)\,(N_1(x, собст) \to \neg\,(N_5(x, отвл) \vee N_5(x, конкр)))$ or
$(\forall x)\,((N_5(x, отвл) \vee N_5(x, конкр)) \to N_1(x, нар))$. These formulas mean that abstract and concrete nouns are nominal ones.

Several examples of the formulas are given below, which may be achieved with the help of similar predicates.

Denote by $Adj_2(x, y)$ a predicate which characterizes a number category of an adjective: y = "ед" ("singular") if x is an adjective in a singular form, y = "мн" ("plural") if x is an adjective in a plural form. And denote by $Adj_3(x, y)$ a predicate for definition of a gender of adjectives: y = "мр" ("masculine") if x is an adjective of masculine gender, y = "жр" ("feminine") if x is an adjective of feminine gender, y = "ср" ("neuter") if x is an adjective of neuter gender.

Than the statement "if an adjective is staying in plural then it is impossible to define its gender" may be written as a formula:

$$(\forall x)\,(Adj_2(x, мр)) \leftrightarrow (\neg Adj_3(x, мр)\&\neg Adj_3(x, жср)\&\neg Adj_3(x, ср))\,.$$

The following formula means the same:

$$(\forall x)\,(Adj_2(x, мн)) \leftrightarrow \neg\,(Adj_3(x, мр) \vee Adj_3(x, жср) \vee Adj_3(x, ср))\,.$$

The statement "if an adjective is in singular then it is obligatory masculine, feminine or neuter, and vice versa if an adjective is masculine, feminine or neuter then it is in singular" may be written by the following formula:

$$(\forall x)\,(Adj_2(x, ед) \leftrightarrow (Adj_3(x, мр) \vee Adj_3(x, жср) \vee Adj_3(x, ср)))\,.$$

Another example. Denote by $PartP_4(x, y)$ a predicate for a definition of aspect of participle: y = "нсв" ("imperfective") if x is a participle of imperfective aspect, y = "св" ("perfective") if x is a participle of perfective aspect.

Table 1. Participle formation

	the present tense	**the past tense**
active voice	imperfective aspect	any
passive voice	imperf.aspect, transitive	transitive

From the table 1, we obtain the following formulas:
$(\forall x)\,((PartP_1(x, cmp)\&PartP_2(x, нcm)) \leftrightarrow (PartP_3(x, ne)\&PartP_4(x, нcв)))$,
which means that passive participles of present tense are formed only from transitive verbs of imperfective aspect;
$(\forall x)\,((PartP_1(x, дcm)\&PartP_2(x, нcm)) \rightarrow PartP_4(x, нcв))$, which means that active participles of pre-sent tense are formed only from verbs of imperfective aspect;
$(\forall x)\,((PartP_1(x, cmp)\&PartP_2(x, npш)) \rightarrow PartP_3(x, ne))$, which means that passive participles of past tense are formed only from transitive verbs;
$(\forall x)\,(PartP_4(x, cв) \rightarrow PartP_2(x, npш))$, which means that participles of perfective aspect can be only in past tense;
$(\forall x)\,(PartP_4(x, нcв) \rightarrow (PartP_2(x, npш) \vee PartP_2(x, нcm)))$, which means that participles of imperfective aspect can be in present tense and in past tense.

The last two formulas may be written in a different way:
$(\forall x)\,(PartP_2(x,) \rightarrow PartP_4(x,))$, which means that participles of present tense can be only of imperfective aspect;
$(\forall x)\,(PartP_2(x, npш) \rightarrow (PartP_4(x, cв) \vee PartP_4(x, нcв)))$, which means that participles of past tense can be of perfective or imperfective aspect.

5. Predicates associated with sentence parts

Furthermore, one can introduce **predicates associated with sentence parts**. Unary predicates of the sentence parts: $P_{sub}(x)$, where x is a subject; $P_{pred}(x)$, where x is a predicate; $P_{adv}(x)$, where x is an adverbial modifier; $P_{attr}(x)$, where x is an attribute; $P_{obj}(x)$, where x is an abject.

Binary predicates of the sentence parts: $P_{sub}(x, y)$, where x is a subject; $P_{pred}(x, y)$, where x is a predicate; $P_{adv}(x, y)$, where x is an adverbial modifier; $P_{attr}(x, y)$, where x is an attribute; $P_{obj}(x, y)$, where x is an abject; where y has a role of a determined word or a word-combination.

For designation of homogeneous parts of sentence (i.e. those parts of sentence which are related to one word and answer one and the same question) we introduce predicates $P_{homo}(x_1, ..., x_n)$ where $x_1, ..., x_n$ are homogenous parts and $P_{homo}(x_1, ..., x_n, y)$, where y is a word or a word-combination, which $x_1, ..., x_n$ are related to.

It is possible to achieve formula representation of these predicates considering x, y as words or word-combinations. Upper index of Q in brackets

is a predicate arity (quantity of predicate places), lower index of Q shows from what part of the sentence we ask a question.

1. The determined word is a subject
$(\forall x, y)\left(Q_1^{(2)}(x,y) \leftrightarrow (P_{sub}(x,y)\&P_{sub}(x)\&P_{pred}(y))\right)$. This formula means that it is possible to raise a question from a subject to a predicate.
$(\forall x, y)\left(Q_1^{(2)}(x,y) \leftrightarrow (P_{obj}(y,x)\&P_{sub}(x)\&P_{obj}(y))\right)$. This formula means that it is possible to raise a question from a subject to an object.

Analogously a formula can be written which means that it is possible to ask question from a subject to an attribute.

2. The determined word is a predicate
$(\forall x, y)\left(Q_2^{(2)}(x,y) \leftrightarrow (P_{adv}(y,x)\&P_{pred}(x)\&P_{adv}(y))\right)$. This formula means that it is possible to raise a question from a predicate to an adverbial modifier.

Formulas which mean that it is possible to raise a question from a predicate to a subject, an attribute, an object, etc. are written analogously.

Besides binary predicates we can introduce many-placed predicates. It is possible if in a sentence one can raise questions from a part of a sentence to some identical parts of a sentence, and the last ones must be not homogeneous (i.e. they must answer different questions or characterize an object or an action from different sides). Since it is known from the syntax foundations that in a simple sentence there cannot be several heterogeneous subjects or predicates at the same time, then only sentences with heterogeneous objects, attributes or adverbial modifiers rest for realization of this case. For example, a formula with a three-place predicate is true for the following sentence:

купить машину нам не по средствам.
$(\forall x, y_1, y_2)\left(Q_2^{(3)}(x,y_1,y_2) \leftrightarrow (P_{obj}(y_1,x)\&P_{obj}(y_2,x)\&P_{pred}(x)\&P_{obj}(y_1)\&P_{obj}(y_2))\right)$ — if $x =$ "купить", $y_1 =$ "машину", $y_2 =$ "нам".

In a general form, formulas for n of heterogeneous sentence parts may be written as follows:
$(\forall x, y_1, ..., y_n)\left(Q_1^{(n+1)}(x,y_1,...,y_n) \leftrightarrow \left(\underset{i=1}{\overset{n}{\&}} P_{attr}(y_i,x)\&P_{sub}(x)\&\underset{i=1}{\overset{n}{\&}} P_{attr}(y_i)\right)\right)$.
This formula means that there are heterogeneous attributes $y_1, ..., y_n$ at a subject x in the sentence.
$(\forall x, y_1, ..., y_n)\left(Q_2^{(n+1)}(x,y_1,...,y_n) \leftrightarrow \left(\underset{i=1}{\overset{n}{\&}} P_{adv}(y_i,x)\&P_{pred}(x)\&\underset{i=1}{\overset{n}{\&}} P_{adv}(y_i)\right)\right)$.
This formula means that there are heterogeneous adverbial modifiers $y_1, ..., y_n$ at a predicate x in the sentence etc.

Several examples of sentences in the form of predicates are presented below.

I. Купить машину нам не по средствам.
$N(машину)$, $ProN(нам)$, $N(средствам)$, $V(купить)$, $Prep(по)$, $PartL(не)$;
$P_{pred}(купить)$, $P_{obj}(машину)$, $P_{obj}(нам)$, $P_{adv}(не\ по\ средствам)$, $P_{adv}(не\ по\ средствам,\ купить)$, $P_{obj}(машину,\ купить)$, $P_{obj}(нам,\ купить)$;
1. $(\forall x, y)\left(Q_2^{(2)}(x,y) \leftrightarrow (P_{adv}(y,x)\&P_{pred}(x)\&P_{adv}(y))\right)$ — if $x =$ “купить”, $y =$ “не по средствам”;
2. $(\forall x, y_1, y_2)\left(Q_2^{(3)}(x,y_1,y_2) \leftrightarrow (P_{obj}(y_1,x)\&P_{obj}(y_2,x)\&P_{pred}(x)\&P_{obj}(y_1)\&P_{obj}(y_2))\right)$ — if $x =$ “купить”, $y_1 =$ “машину”, $y_2 =$ “нам”.

II. Она шла нетвердой походкой.
$ProN(она)$, $V(шла)$, $N(походкой)$, $Adj(нетвердой)$;
$P_{sub}(она)$, $P_{pred}(шла)$, $P_{attr}(нетвердой)$, $P_{adv}(походкой)$, $P_{sub}(она,\ шла)$, $P_{adv}(походкой,\ шла)$, $P_{attr}(нетвердой,\ походкой)$, $P_{pred}(шла,\ она)$;
1. $(\forall x, y)\left(Q_1^{(2)}(x,y) \leftrightarrow (P_{sub}(x,y)\&P_{sub}(x)\&P_{pred}(y))\right)$ — if $x =$ “она”, $y =$ “шла”;
2. $(\forall x, y)\left(Q_2^{(2)}(x,y) \leftrightarrow (P_{pred}(x,y)\&P_{pred}(x)\&P_{sub}(y))\right)$ — if $x =$ “шла”, $y =$ “она”;
3. $(\forall x, y)\left(Q_2^{(2)}(x,y) \leftrightarrow (P_{adv}(y,x)\&P_{pred}(x)\&P_{adv}(y))\right)$ — if $x =$ “шла”, $y =$ “походкой”;
4. $(\forall x, y)\left(Q_4^{(2)}(x,y) \leftrightarrow (P_{attr}(y,x)\&P_{adv}(x)\&P_{attr}(y))\right)$ — if $x =$ “походкой”, $y =$ “нетвердой”.

III. Самолет, пролетавший над нами, скрылся в облаках.
$N(самолет)$, $N(облаках)$, $ProN(нами)$, $V(скрылся)$, $PartP(пролетавший)$, $Prep(над)$, $Prep(в)$;
$P_{sub}(самолет)$, $P_{pred}(скрылся)$, $P_{attr}(пролетавший\ над\ нами)$, $P_{adv}(в\ облаках)$, $P_{adv}(в\ облаках,\ скрылся)$, $P_{sub}(самолет,\ скрылся)$, $P_{pred}(скрылся,\ самолет)$, $P_{attr}(пролетавший\ над\ нами,\ самолет)$;
see II.1. – if $x =$ “самолет”, $y =$ “скрылся”;
$(\forall x, y)\left(Q_1^{(2)}(x,y) \leftrightarrow (P_{attr}(y,x)\&P_{sub}(x)\&P_{attr}(y))\right)$ — if $x =$ “самолет”, $y =$ “пролетавший над нами”;
see II.2. — if $x =$ “скрылся”, $y =$ “самолет”;
see II.3. — if $x =$ “скрылся”, $y =$ “в облаках”.

IV. Ни Пети, ни Тани в школе в тот день не было.
$N(Пети)$, $N(Тани)$, $N(школе)$, $N(день)$, $V(было)$, $Adj(тот)$, $Prep(в)$, $Con(ни)$, $PartL(не)$;
$P_{attr}(тот)$, $P_{pred}(не\ было)$, $P_{obj}(Тани)$, $P_{obj}(Пети)$, $P_{adv}(в\ школе)$, $P_{adv}(в\ день)$, $P_{obj}(Тани, не\ было)$, $P_{obj}(дети, не\ было)$, $P_{adv}(в\ школе, не\ было)$, $P_{adv}(в\ день, не\ было)$, $P_{attr}(тот, день)$,

$P_{homo}(Тани, Пети)$, $P_{homo}(Тани, Пети, не\ было)$.
1. $(\forall x, y)\left(Q_2^{(2)}(x, y) \leftrightarrow (P_{obj}(y, x)\&P_{pred}(x)\&P_{obj}(y))\right)$ — if $x =$ "не было", $y =$ "Тани" or $y =$ "Пети";
2. $(\forall x, y)\left(Q_2^{(2)}(x, y) \leftrightarrow (P_{adv}(y, x)\&P_{pred}(x)\&P_{adv}(y))\right)$ — if $x =$ "не было", $y =$ "в день" or $y =$ "в школе";
3. $(\forall x, y)\left(Q_4^{(2)}(x, y) \leftrightarrow (P_{attr}(y, x)\&P_{adv}(x)\&P_{attr}(y))\right)$ — if $x =$ "в день", $y =$ "тот".

6. Syntactic valences of words (verbs)

As an intermediate result we obtain that it is possible to define syntactical valences of a word with the help of in predicates mentioned above.

We have the following facts from the syntax theory [7], [8]:

Compound verbal predicate consists of auxiliary verb and infinitive, i.e. substantial (main) part. *Compound noun predicate* consists of copula and noun part, which is expressed by noun, adjective, participle, numeral, pronoun, adverb or interjection. *Compound predicate* is a predicate, which consists of three or more words and, as a rule, has characteristics of compound verbal predicate and compound noun predicate.

Defining a syntactic valency of a verb in a sentence we have three situations. The first one is when a verb in a sentence is a part of simple verbal predicate or compound noun predicate. If predicates are homogeneous we define a valency of only one of them. Valences of the other verbs will be the same. It is easy to define valences and corresponding actants of such verb by quantity of the questions which can be asked from it.

The second situation is when a verb is a part of a compound predicate or compound verbal predicate. The first part of it (auxiliary verb and copula) is connected with a subject and the second one (infinitive or noun part) has a connection with the other parts of sentence. In other words, an auxiliary verb has a valency equal to 1 (if a sentence is not impersonal) and a substantial part of compound verbal predicate has a valency, which is equal to number of questions, which can be asked from a subject. In terms of predicates which are associated with parts of a sentence, as was introduced before. It means that a valency of a main part of compound verbal predicate is equal to number of predicates defined in this sentence, on the second place of which there is a predicate and on the first one stands not a subject. In a compound predicate a valency of an auxiliary verb is defined as in a compound verbal predicate. Thereby the number of secondary parts of a sentence connected with a compound predicate is a valency of the last verb of a substantial part.

The third situation is when a verb is a secondary part of a sentence. In this case for defining a valency and actants we do the same as in the first case. But in the third situation a valency of a verb exceeds 1 very rarely

(if it is possible at all). Thus, with the help of introduced predicates, which correspond to parts of a sentence, it is possible to determine a syntactic valency of verbs in a sentence.

7. Henkin's construction and sense of text

Thus, we have a set of various predicates and formulas of first order logic matching sentences in a natural language, based on the grammatical, syntactic and semantic structures of words and sentences.

We repeat that our concept consists of using Henkin's construction from mathematical logic to construct some finite models, which can be treated as a sense of a text [9].

However this construction should be modified. Certainly, it is possible to consider elements of the dictionary of a natural language as constants; to enter predicates by any ways described above, to receive formulas on the base of them. And certainly a specialist can easyly understand what it means, but some questions must be separately considered.

At application of Henkin's construction it can appear, that the theory has no finite models, only infinite ones. In this case in practice, we simply interrupt process of construction of model at any stage. If we consider not mathematical, but usual "household" texts in a natural language, taking such texts as a basis, we will discover that, theories without quantifiers, and in particular, without free variables will appear, i.e. there will be only constants, and corresponding models will be finite.

There is one more question, which consist of the problem that theories arising from the text in a natural language can be unsolvable. But really we can verify only partial consistency. For example, if a subject is "white", it cannot be "black". The symmetry or transitivity of some predicates can be checked up. For example, the predicate "to be relatives" is symmetric, and predicates "above" or "under" are transitive.

Resuming, it is possible to say, that we need to use not accurate Henkin's construction, but some approximated analogue. Note separately, that a procedure of construction of mentioned above finite models, which can be treated as sense of the text, with corresponding assumptions, becomes a formal procedure. Such procedure can be realized on the computer.

Now let us consider some technical questions more detailed. Suppose we have a text, not a separate sentence.

Thus there is a text, i.e. final set of sentences, $p_1 p_2 ... p_N$, at the input. Some streams are formed at the output:

$S_1 = \langle s_{11}, s_{12}, ..., s_{1m_1}, ... \rangle$

. . . . . .

$S_k = \langle s_{k1}, s_{k2}, ..., s_{km_k}, ... \rangle$

An elementary auxiliary stream consists of ordered pairs $\langle 1, p_1, 2, p_2, ..., N, p_N \rangle$, where the first component is the sentence number, the second one is the sentence itself.

Information about word-formation may be placed in streams like $\langle h, k_1, L_1, k_2, L_2, ... \rangle$, where h is the stream heading, for instance a selected suffix; k_i is the sentence number, where the word with this suffix appears (i.e. k_i are numbers not for all sentences but only for those, where these words appear); L_i is the list of words with the given suffix appearing in the sentence.

Streams may be associated with lexical functions too. Finite models matching with source text will also represented in the form of streams.

For instance, let us select all nouns from sentences and write them in a stream $\langle 1, n_1^1, ..., n_{l_1}^1; 2, n_1^2, ..., n_{l_2}^2; ... \rangle$, where sentence numbers and lists of nouns entering in this sentence are written in series (l_i is a list size). We rewrite this stream in the form $\langle \langle 1, n_1^1 \rangle, ..., \langle 1, n_{l_1}^1 \rangle, \langle 2, n_1^2 \rangle, ..., \langle 2, n_{l_2}^2 \rangle, ... \rangle$.

Denote by $C = \{\langle t, n_j^t \rangle | t = \overline{1, N}, j = \overline{1, l_t}\}$ a set of all pairs that appear in the stream. The basic sets of models will be sets of a form $C_0/\sim$, where $C_0 \subseteq C$, $\sim$ is some equivalence relation.

Equivalence relations will appear almost in the same way as they appear in Henkin's construction in the proving process of the model existence theorem[6], i.e. pairs of the type, $\langle t, c_j^t \rangle$ $(t = 1, ..., N)$ may be considered as constants depending on various propositions about these constants which we regard equivalent.

Analogously using the obtained stream it will be possible to apply the types omitting theorem[6], and additionally get some models as a result.

Note that in the process of using Henkin's constructions it is essential to check a consistency of corresponding theories at every stage. However only a partial testing of a noncontradictory can be used in a computer processing of a natural language text. For example we check that relations as "over" or "under" are really transitive; if it is said "white" about an object, then it is not "black" and so on.

8. Conclusion

Different approaches to representing semantics of natural language texts are of great interest now. That is why we have made efforts to analyze the sense of the text on basis of structural analysis of sentences and the text as a whole.

Large amount of predicates and logic formulas of the first order were proposed for such analysis. However we note that in the main the given predicates and formulas are concerned with a grammatical and syntactic structure of sentences.

In the future the achieved results may be studied and transformed by mathematical logic methods. It gives us an opportunity to make a transition from a syntactic to a semantic level.

This work can be used for creation of a text sense theory, and it is possible to apply the results of this work in a mathematical logic area and in linguistic investigations.

Thus in spite of the fact that this stage of the work is absolutely necessary, it is important to note that in the current time semantic text structure has not been reflected adequately in achieved formulas, and additional investigations are necessary.

We also note that the large volume of factual information from classical and mathematical linguistics, and mathematical logic was used at this simplest (in our opinion) stage. It tells us about difficulty of this problem in the whole.

References

[1] Apresyan U.D. Experimental Semantic Invesigation of a Russian Verb. — M.: Nauka, 1967. — 251 p. (in Russian)

[2] Melchuk I.A. Experience of Theory of Linguistic Models like "Sence <-> Text". — M.: Nauka, 1974. — 315 p. (in Russian)

[3] Markus S. Set-theoretical Models of Languages. — M.: Nauka, 1970. — 332 p. (in Russian)

[4] Sacks G.E. Saturated Model Theory. — M.: Mir, 1976. — 192 p. (in Russian, translated from W.A. Benjamin Inc. 1972)

[5] Paducheva E.V. Dynamic models in semantics of lexic. — M.: Languages of Slavic culture, 2004. — 608 p. (in Russian)

[6] Lakhuti, D.G.: The automatic analysis of texts in a natural language // Information Processes and Systems. — Moscow, 2003. — N 11. — P. 18–25. (in Russian)

[7] Beloshapkova V.A. Modern Russian Language: Manual for Philology Students of Institutes of Higher Education. — M.: Azbukovnik, 1997. — 928 p. (in Russian)

[8] Rosental D.E. Modern Russian Language: Manual for Philology Students of Institutes of Higher Education. — M.: MSU Edition, 1971. — 636 p. (in Russian)

[9] Batura T., Murzin F. Logical Methods for Representing Meaning of Natural Language Texts // Proc. 4th Internat. Conf. on Computational Science — ICCS 2004, Krakow, Poland, June 6-9, 2004. Part 3. — Lect Notes Comput. Sci. — 2004. — Vol. 3038. — P. 545–551.