Processing and understanding of the natural language by an intelligent system

Hardzei A.
Minsk State
Linguistic University
Minsk, Belarus
Email: alieks2001@yahoo.com

Svyatoshchik M.
Minsk State
Linguistic University
Minsk, Belarus
Email: svyatoshchikm@mail.ru

Bobyor L.
Intelligent Semantic Systems Ltd.
Minsk, Belarus
Email: lizabobyor@gmail.com

Nikiforov S.
Belarusian State University
of Informatics and Radioelectronics
Minsk, Belarus
Email: nikiforov.sergei.al@gmail.com

***Abstract*—The article is dedicated to Natural Language Processing in the Theory for Automatic Generation of Knowledge Architecture (TAPAZ-2) paradigm and the immersion of the obtained semantic formalisms into the software environment through the Open Semantic Technology for Intelligent Systems (OSTIS). A specific feature of the approach is the formalization of natural language semantics based on the World Model and the combination of Semantic Coding with the ontology and taxonomy of semantic networks.**

***Keywords*—Natural Language Processing (NLP), Natural Language Understanding (NLU), macroprocess, roles of individs, Parts of Language, Parts of the Sentence, taigens, yogens, Combinatory Semantics, Theory for Automatic Generation of Knowledge Architecture (TAPAZ-2), ontology, taxonomy, semantic network, Open Semantic Technology for Intelligent Systems (OSTIS), SC-code (Semantic Computer Code)**

I. Introduction

The present research is carried out in the framework of Combinatory Semantics, which studies the linguistic mapping of the dynamics of individs' roles in an event. [1, p. 13]. Natural language is understood as a system of figures and signs for decoding the World Model and conscious management of intellectual activity [2, p. 35]. The World Model (hidden knowledge) is the architecture of patterns, i.e. the ordered set of patterns and the ordered set of transformations of some patterns in others [3, p. 226], [46, p. 182]. It is necessary to distinguish between verbal and non-verbal knowledge. Non-verbal knowledge is beyond the rational approach for comprehension of the World, it is beyond any term system, whether it is mathematics, computer science, linguistics, paralinguistics and semiotics as a whole. It is impossible to explain and show, how the imageries of Raphael's paintings came up, therefore, non-verbal knowledge should not be confused with facial expressions and gestures. Sign or finger language (language of the deaf and dumb) is as a natural language as any other hieroglyphic language [4, p. 18]. Verbal knowledge consists of information and fascination. According to Yu. V. Knorozov, the maximal information is contained in mathematics and the maximal fascination is contained in music [5]. Further studies have shown that fascination, along with factuation or factualization, is still a kind of information [6].

In this case, fascination that involves any stylistic nuance, all kinds of emotional and expressive shades and "induced" emotions, including those created with the help of meter, rhythm, pause, chant, representational devices and other accentological means, are equally covered by declarative and procedural methods of representing knowledge. Due to the fact that language categories as supporting constructs of the metatheory of any natural language are linked to verbal knowledge, only those that are distinguished procedurally, fixed declaratively and confirmed combinatorially can be determined as relevant [7]. Meanwhile, "The "chunks" of reasoning, language, memory, and "perception" ought to be larger and more structured; their factual and procedural contents must be more intimately connected in order to explain the apparent power and speed of mental activities" [8, p. 1].

The purpose of the article is to demonstrate the possibility of understanding texts in a natural language by computer systems with semantic software that allows creating a problem solver architecture based on a combination of Semantic Coding with the ontology and taxonomy of semantic networks.

To write formalized texts, in the article, the variants of the external displaying of SC-code constructs – SCg (graphic version) and SCn (hypertext version) – are used.

II. Problem definition

The discrepancy of the World Model and the Linguistic Image of the World, indefinite meaning of lexical units and syntactic incompleteness of sentences are the main stumbling blocks in Natural Language Processing (NLP) [9]. Attempts to remove or circumvent these problems with the help of statistics based on co-occurrence by Z. S. Harris [10] resemble guessing the correct answers by schoolchildren during centralized testing. No matter how sophisticated the methods of statistical processing of structured or unstructured natural language content are, they only imitate the intellectual or inventive activity of a human, guessing the correct answer with more or less certainty, but we do not doubt that neural networks are able to efficiently scale the solutions found by combinatory methods. As for the currently popular combinatory methods, they go back to the semantic cases of Ch. Fillmore [11]–[22] and Stowell's "theta-grids" [23] and are used, in particular, in the Semantic Web project of T. Berners-Lee [24]–[27] and on an international community resource Schema.org of the Google, Microsoft, Yahoo and Yandex developer communities [28]. The main disadvantage of these methods is their empirical character and the lack of unified algebraic bases of semantic calculus. Because of these reasons, the creators of the Semantic Web, despite titanic efforts to standardize the technology, have not yet managed to reduce various subject ontologies to a top-level ontology, which, as many commentators emphasize, is "critical for the entire concept" [29, p. 94]. The fact is that the top-level ontology cannot be built from below, it, so to speak, "does not grow" from the ontologies of subject domains but must be initially set from above and in an algebraic standard that is suitable for formalizing texts in natural language including sentences, free and stable strings of combinatory variants of lexical items and lexical items themselves that make up these texts. In other words, to embed patterns of the World Model in presentation formalisms, a formalized language is required, which is comparable in semantic power to a natural language, as V. V. Martynov pointed out at the time [30]. Otherwise, as a result, we will get, as D. G. Bobrow and T. Winograd wrote, an extremely fragile structure, often resembles a house of cards and collapses immediately "if swayed in the slightest from the specific domain (often even the specific examples) for which it was built" [31, p. 4].

Let us pay attention one more time: semantics as the relation of language to the World Model is manifested in the dynamics of individs' roles in an event, which is reflected in the content of patterns, the meaning of signs and the sense of sentences [1]–[4]. It is possible to arbitrarily declare any top-level object-oriented programming language, such as Java, C++, C# or the next version of the OWL language from the Semantic Web project, as a top-level ontology, but until such languages can encode the content of patterns, the meaning of signs and the sense of sentences and then reduce them to semantic primitives underlying calculus [32], such statements will be only declarations. If the OWL language allowed encoding patterns of the World Model and conjugate code with natural language semantics, then the Internet would already be turned into the Global Artificial Intelligence through the Semantic Web project. It should be noted that linguistics has only one synthetic (consistently deductive and procedural) language model – Panini grammar that is dated from the 5th century BC, in which, with the help of 3959 short rules (sutras), the generation, construction and transformation of all Sanskrit units are exhaustively described, starting from the phonetic-phonological level and ending with the semantic-syntactic one [33]–[35]. Unfortunately, it has not yet been clarified what formalisms are the basis for such an accurate description of natural language and how it was possible to achieve this in such a long time. From modern methods of encoding language semantics, six versions of Universal Semantic Code (USC) of V. V. Martynov [30], [36]–[40] and their finalization in the Theory for Automatic Generation of Knowledge Architecture (TAPAZ-2) by A. N. Hardzei [4], [41]–[45] are known.

TAPAZ Semantic Classifier as a top-level ontology includes the Set of Macroprocesses as Semantic Primitives (Paradigm of Actions) ordered by TAPAZ-algebra, Role List of Individs and TAPAZ Knowledge Graph [46].

Taking into account that the calculus of subject domains and the semantics of each subject domain is implemented in TAPAZ-2 separately using a specially oriented knowledge graph, the most effective means of immersing TAPAZ formalisms in the software environment are dynamic graph models, primarily an SC-code (Semantic Computer Code) of the Open Semantic Technology for Intelligent Systems (OSTIS) developed by the school of V. V. Golenkov [47], [70], [71].

We suppose that combining efforts and an organic combination of semantic coding with the ontology and taxonomy of semantic networks will solve a number of central problems of automatic data processing in natural language (Natural Language Processing), shifting the emphasis towards machine understanding of natural language (Natural Language Understanding, NLU).

III. Proposed approach

According to T. N. Domnina and O. A. Khachko, at the beginning of 2014, the number of scientific peer-reviewed journals was 34,274. If the average amount of articles is at least 50 per year, then 1,713,700 articles are published per year [48]. T. V. Chernigovskaya complains that "the number of articles related to the brain exceeds 10 million – they simply cannot be read. Every day about ten pieces get out" [49]. The average growth in the number of peer-reviewed scientific journals is 4% per year. In 2018, 1.6 million scientific articles were included in the Web of Science database [50]. So, it is essential to use automatic text analysis, artificial intelligent systems for searching and processing information.

In 1994, A. N. Hardzei, in a group led by V. V. Martynov, for the first time proposed a procedure for calculating semantics in the form of a specially oriented graph for ranking complex strings [41]. Use of the procedure has required the establishment of a one-to-one (vector) transition between actions in basic semantic classifier and has led to the creation of the Theory for Automatic Generation of Knowledge Architecture (TAPAZ) which was founded on: the formal theory; the semantic counterpart; the set of macroprocesses (actions) as semantic primitives; the algorithm defining roles of individs, and the graph for searching processes through macroprocesses (knowledge graph) [44, p. 11]

In 2014, the second version of TAPAZ appeared, in which greatly simplified algebraic apparatus, increased number of rules of interpretation of the standard superposition of individs [43].

At the same time, the problems of unifying the principles of building various components of computer systems were solved within the framework of the OSTIS project [51] aimed at creating an open semantic technology for engineering knowledge-driven systems. This technology allows combining heterogeneous models of problem solving as a universal platform and reducing costs that arise during development and modification, including when adding new components to the system. The OSTIS Technology makes it possible to use both statistical and combinatory methods that operate with knowledge. It is based on a unified version of information encoding based on semantic networks with a basic set-theoretic interpretation called SC-code. The architecture of each system built using the OSTIS Technology (ostis-system) includes a platform for interpreting semantic models of ostis-systems as well as a semantic model of the ostis-system described using the SC-code (sc-model of the ostis-system). In turn, the sc-model of the ostis-system includes the sc-model of the knowledge base, the sc-model of the problem solver and the sc-model of the interface (in particular, the user one).

The basis of the knowledge base of any ostis-system (sc-model of the knowledge base) is a hierarchical system of subject domains and corresponding ontologies. The upper level of the hierarchy of the knowledge base fragment related directly to natural language processing is shown below.

Knowledge base on natural language processing
⇐ *section decomposition**:
{
- *Section. Subject domain of lexical analysis*
- *Section. Subject domain of syntactic analysis*
- *Section. Subject domain of semantic analysis*
- *Section. Subject domain of TAPAZ-2*

}

The problem solver of any ostis-system (sc-model of the ostis-system problem solver) is a hierarchical system of agents of knowledge processing in semantic memory (sc-agents) that interact with each other solely by specifying the acts they perform in the specified memory.

Problem solver for natural language processing
⇐ *decomposition of an abstract sc-agent**:
{
- *Agent of lexical analysis*
- *Agent of syntactic analysis*
- *Agent of semantic analysis*
- *Agent of merging structures in the knowledge base*
- *Agent of logical inference*

}

Agent of merging structures in the knowledge base
⇐ *decomposition of an abstract sc-agent**:
{
- *Agent of searching for contradictions*
- *Agent of resolving contradictions*

}

The agent of lexical analysis decomposes the text into lexemes and nominative units (stable strings of combinatory variants of lexemes) based on the dictionary included in the subject domain of lexical analysis. Note that the lexicographic description also presupposes the establishment of the linguistic semantic category for the lexeme, i.e. its belonging to a certain Part of Language [52]–[56]. The agent of syntactic analysis builds the syntactic structure of the analyzed text using the specified rules. The agent of semantic analysis performs the transition from the text specification created by the previous agents to the structure that describes its semantics. The agent of merging structures in the knowledge base compares the structures obtained as a result of the text analysis with the data available in the knowledge base and, if contradictions are detected, resolves them.

The agent of logical inference uses logical rules written by means of the SC-code and interacts with the agents of syntactic and semantic analysis.

A more detailed explanation of the abovementioned subject domains and agents of the proposed approach is given on the example of processing of a particular fragment of natural-language text, namely the description of the technological process of production of cottage cheese: «Производство творога кислотным способом включает в себя: приёмку молока, нормализацию молока до жирности 15%, очистку и пастеризацию молока, охлаждение молока до температуры заквашивания, внесение закваски в молоко, сквашивание молока, разрезку сгустка, подогрев и обработку сгустка, отделение сыворотки, охлаждение творога»[1].

[1] "Manufacture of cottage cheese by the acid method includes: acceptance of milk, normalization of milk to 15% fat, purification and pasteurization of milk, cooling of milk to the fermentation temperature, adding sourdough to milk, fermentation of milk, cutting of the clot, heating and processing of the clot, separation of whey, cooling of cottage cheese".

From the point of view of the ostis-system, any text is a file (i.e., an sc-node with content). An example of such a node is shown in Figure 1.

Figure 1. Representation of natural language text in the system.

Let us consider each of the stages of processing this text.

IV. Lexical analysis

It is a decomposition of a text by an agent of lexical analysis into separate lexemes and stable strings of combinatory variants of lexemes (nominative units) based on a dictionary that is part of the subject domain of lexical analysis. Below is a fragment of the ontology that contains knowledge about Parts of Language.

lexeme
$\subset$ *file*

nominative unit
$\subset$ *file*

combinatory variant of the lexeme
$\subset$ *file*

The *lexeme* is a taigen or yogen of a particular natural language [2, p. 35]. A *combinatory variant of lexeme* is a variant of a lexeme in an ordered set of its variants (paradigm) [57, p. 351].

A *nominative unit* is a stable string of combinatory variants of lexemes, in which one variant of the lexeme (modificator) defines another one (actualizator), for example: 'записная книжка' = 'note book', 'бежать галопом' = 'run at a gallop' [2, p. 35].

morphological paradigm*
$\in$ *quasi-binary relation*
$\Rightarrow$ *first domain**:
word form
$\Rightarrow$ *second domain**:
lexeme

natural language
$\Rightarrow$ *decomposition**:
{• *Part of Language*
$\Rightarrow$ *decomposition**:
{• *taigen*
• *yogen*
}
• *sign of syntax alphabet*
}

The *morphological paradigm** is a quasi-binary relation, connecting a lexeme with its combinatory variants. The *lexeme* – taigen or yogen of a particular natural language, being a sign, it has a combination of figures in the aspect of expression, and it has a pattern in the aspect of content; in synthetic languages it has a developed morphological paradigm and is the central unit of lexicographic description.

*Signs of Syntax Alphabet** are auxiliary syntactic means (at the macrolevel – prepositions, postpositions, conjunctions, particles, etc., at the microlevel – flexions, prefixes, postfixes, infixes, etc.) that serve for connecting the components of language structures and the formation of morphological paradigms [2, p. 35].

A *nominative unit* is a stable string of combinatorial variants of lexemes, in which one variant of a lexeme (modificator) defines another (actualizator).

taigen
$\Rightarrow$ *decomposition**:
{• *expanded taigen*
$\Rightarrow$ *decomposition**:
{• *composite taigen*
• *complex taigen*
}
• *reduced taigen*
$\Rightarrow$ *decomposition**:
{• *contracted taigen*
• *constricted taigen*
}
}

constricted taigen
$\Rightarrow$ *decomposition**:
{• *informational taigen*
• *physical taigen*
$\Rightarrow$ *decomposition**:
{• *constant taigen*
• *variable taigen*
}
$\Rightarrow$ *decomposition**:
{• *quantitative taigen*
• *qualitative taigen*
}
$\Rightarrow$ *decomposition**:
{• *single-place taigen*
• *multi-place taigen*
$\supset$ *intensive taigen*
$\supset$ *extensive taigen*
}
}

A *taigen* is a Part of Language that denotes an individ.

An *informational taigen* denotes an individ in the informational fragment of the World Model, a *physical taigen* denotes an individ in the physical fragment of the World Model.

A *constant taigen* denotes a constant individ, a *variable taigen* denotes a variable individ [54, pp. 70–72], [58, pp. 12–13].

yogen
⇒ *decomposition**:
{• *expanded yogen*
  ⇒ *decomposition**:
  {• *composite yogen*
  • *complex yogen*
  }
• *reduced yogen*
  ⇒ *decomposition**:
  {• *contracted yogen*
  • *constricted yogen*
  }
}

contracted yogen
⇒ *decomposition**:
{• *informational yogen*
• *physical yogen*
  ⇒ *decomposition**:
  {• *constant yogen*
  • *variable yogen*
  }
  ⇒ *decomposition**:
  {• *quantitative yogen*
  • *qualitative yogen*
  }
  ⇒ *decomposition**:
  {• *single-place yogen*
  • *multi-place yogen*
  }
}

multi-place yogen
⇒ *decomposition**:
{• *intensive yogen*
• *extensive yogen*
}

A *yogen* is a Part of Language that denotes an attribute of an individ.

An *informational yogen* denotes the attribute of an individ in the informational fragment of the World Model, a *physical yogen* denotes the attribute of an individ in the physical fragment of the World Model.

A *constant yogen* denotes a constant attribute of an individ, a *variable yogen* denotes a variable attribute of an individ [54, pp. 71–74], [58, p. 12–13].

The lexemes in the knowledge base are described in the form shown in Figure 2.

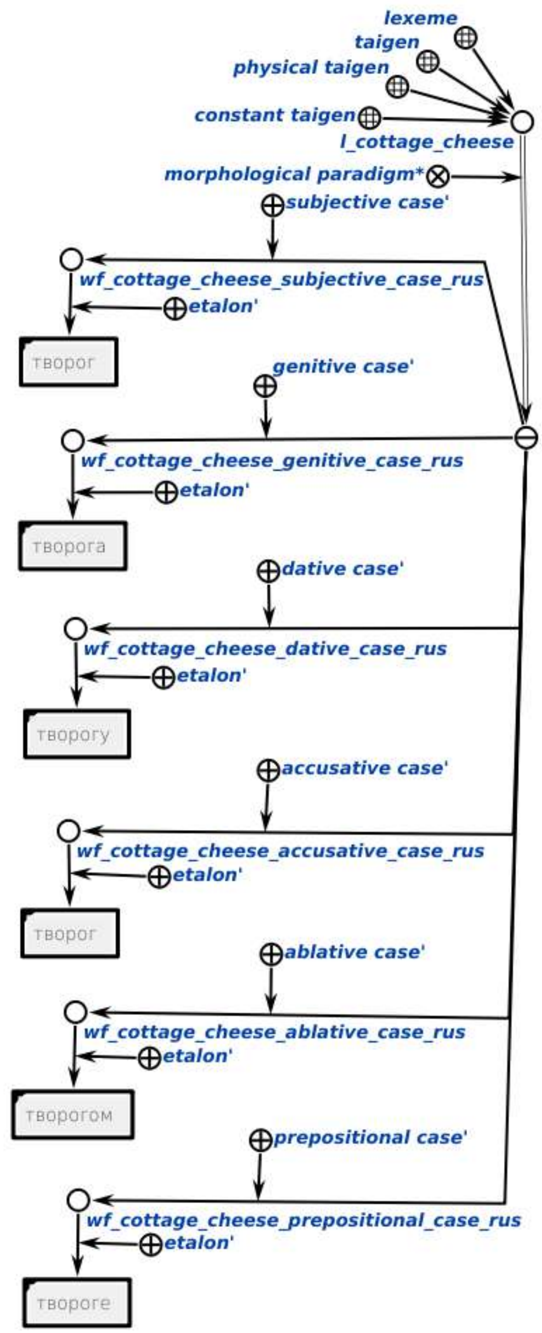

Figure 2. The description of the lexeme in the knowledge base.

The construct, which is the result of lexical analysis, is shown in Figure 3.

V. Syntactic analysis

The agent of the syntactic analysis performs the transition from the lexically marked text to its syntactic structure based on the rules described in the correspond-

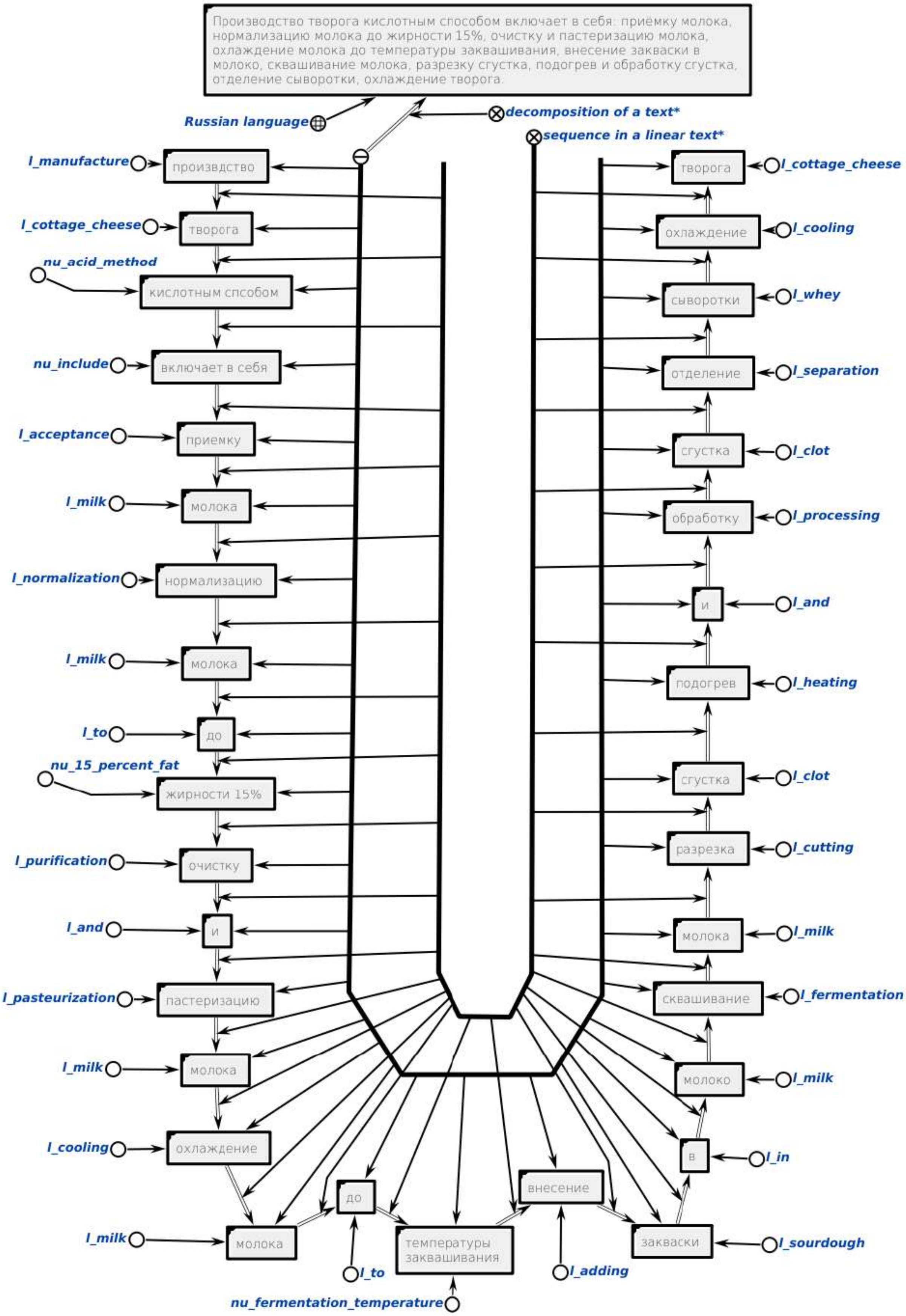

Figure 3. The result of lexical text analysis.

ing subject domain. A fragment of the ontology of the subject domain is presented below:

part of the sentence'
∈ *role relation*
⇒ *decomposition**:
- { *principal part of the sentence'*
  - ⇒ *decomposition**:
    - { *grammatical subject'*
    - *grammatical predicate'*
    - *grammatical direct object'*
    - }
- *subordinate part of the sentence'*
  - ⇒ *decomposition**:
    - { *grammatical indirect object'*
    - *grammatical attribute'*
    - *grammatical circumstance'*
    - }
- }

A *Part of the Sentence'* is a relation that connects the decomposition of a text with a file whose contents (Part of Language) play a certain syntactic role in the decomposed text [2, p. 35].

The *grammatical subject'* is one of the principal role relations that connects the decomposition of a text with the file, the contents of which denotes the starting point of the event description selected by the observer.

The *grammatical direct object'* is one of the principal role relations that connects the decomposition of a text with the file, the contents of which denotes the final point of the event description selected by the observer.

The *grammatical predicate'* is one of the principal role relations that connects the decomposition of the text with the file, the contents of which denotes the mapping by the observer of the starting point of the event description to the final point [46, p. 184].

A *grammatical circumstance'* is one of the subordinate role relations that connects the decomposition of a text with a file, the contents of which denote either a modification or localization of the grammatical predicate; the grammatical circumstance of degree and the grammatical circumstance of manner denote the modification of the grammatical predicate, the grammatical circumstance of place and the grammatical circumstance of time denote the spatial and, accordingly, the temporal localization of the grammatical predicate.

A *grammatical attribute'* is one of the subordinate role relations that connects the decomposition of a text with the file, the contents of which denote a modification of the grammatical subject, grammatical object, grammatical circumstance of place and time [57, pp. 352–354, 357], [59, pp. 29–33], [60].

grammatical circumstance'
⇒ *decomposition**:
- { *grammatical circumstance of degree'*
- *grammatical circumstance of manner'*
- *grammatical circumstance of place'*
  - ⇒ *decomposition**:
    - { *dynamical grammatical circumstance of place'*
    - *static grammatical circumstance of place'*
    - }
- *grammatical circumstance of time'*
  - ⇒ *decomposition**:
    - { *dynamical grammatical circumstance of time'*
    - *static grammatical circumstance of time'*
    - }
- }

A fragment of the ontology that is the result of this stage is presented in Figures 4 and 5.

VI. Construction in terms of TAPAZ-2

The agent of semantic analysis performs the transition from the processed text to its semantics formulated in terms of TAPAZ-2 on the basis of the rules described in the corresponding subject domain [4], [42]–[46], [61]. A fragment of the ontology of this subject domain is presented below:

participant of the exposure*
:= [participant of the event*]
∈ *non-role relation*
⇒ *first domain**:
individ
⇒ *second domain**:
action
⇒ *decomposition**:
- { *subject**
  - ⇒ *decomposition**:
    - { *initiator**
    - *inspirer**
    - *spreader**
    - *creator**
    - }
- *instrument**
  - ⇒ *decomposition**:
    - { *activator**
    - *suppressor**
    - *enhancer**
    - *converter**
    - }
- *mediator**
  - ⇒ *decomposition**:
    - { *landmark**
    - *locus**
    - *carrier**
    - *adapter**
    - *acceptor**

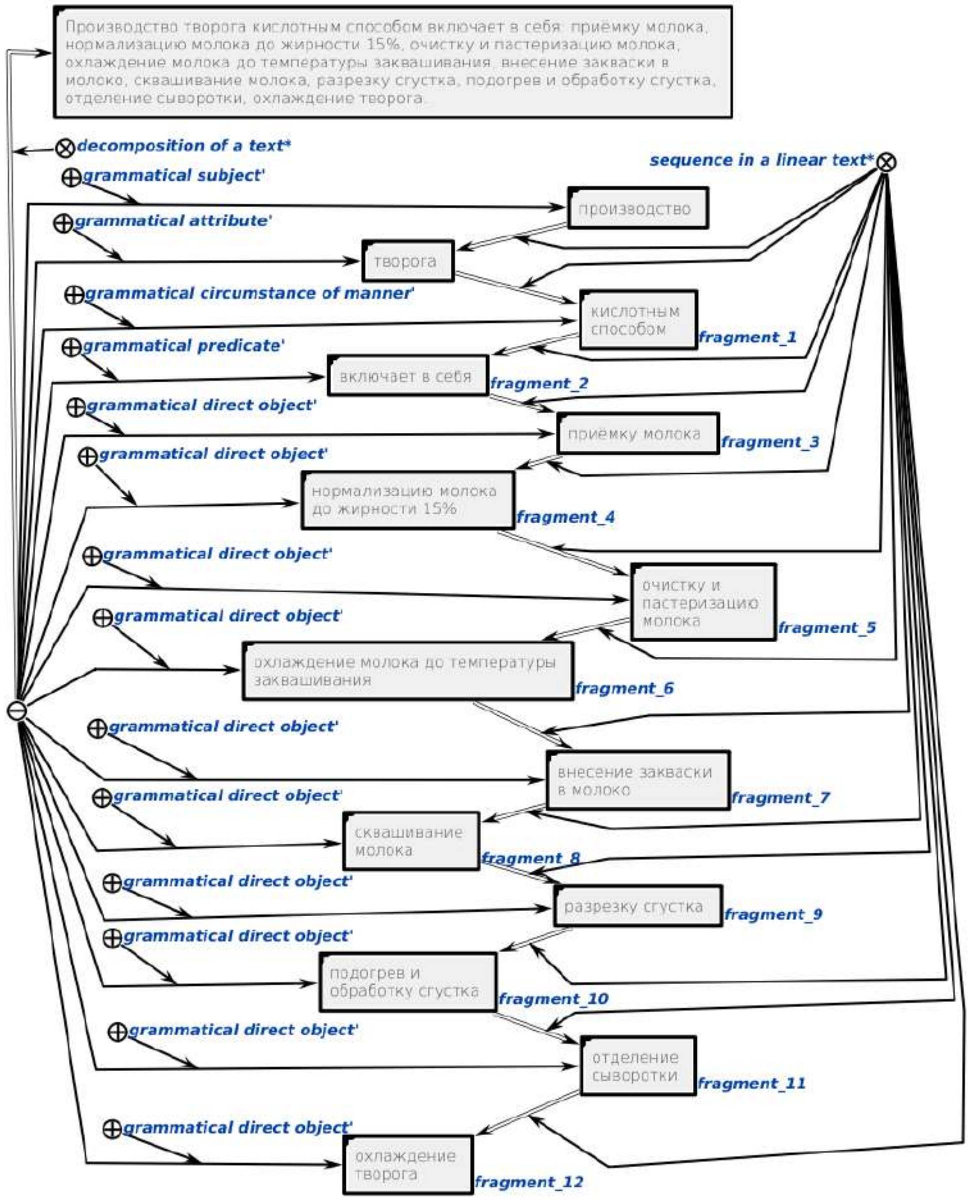

Figure 4. The result of the syntactic analysis of the text, the first fragment.

- *stock**
- *separator**
- *material**
- *model**
- *retainer**
- *resource**
- *stimulus**
- *regulator**
- *chronotope**
- *source**
- *indicator**

}

- *object**

⇒ *decomposition**:

{
- *coating**
- *hull**
- *interlayer**
- *kernel**

}

- *product**

⇒ *decomposition**:

{
- *billet**
- *semi-product**
- *prototype**
- *end item**

}

}

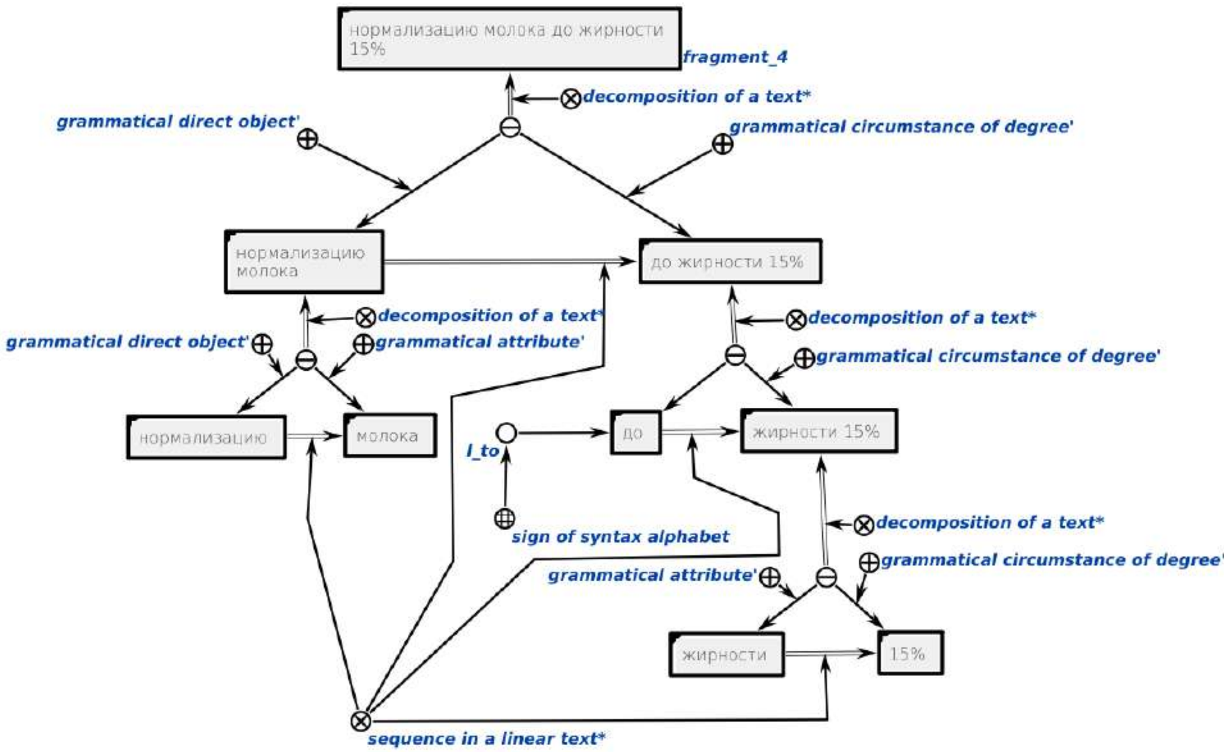

Figure 5. The result of the syntactic analysis of the text, the second fragment.

The *individ* is a kind of the pattern as a separate entity in the selected fragment of the World Model [2, p. 34].

The *participant of the action** is a non-role relation that connects the action with the individ that participates in it.

The *subject** – the originator of the action, varieties of the subject: *initiator** – initiates the action, *spreader** – spreads the action, *inspirer** – involves into the action, *creator** – completes the action by making a product from the object.

*Instrument** – the performer of the action, the closest individ to the subject, varieties of the instrument: *activator** – directly affects the mediator, *suppressor** – suppresses the resistance of the mediator, *enhancer** – increases the effect on the mediator, *converter** – converts the mediator into the instrument.

*Mediator**, i.e. the mediator of the action – the closest individ to the object; varieties of the mediator: *landmark** – orientates the impact on the object, *locus** – the closest environs of the object partially or completely surrounding the object that localizes the object in space and thereby containing (enclosing) it, *carrier** – carries the object, *adapter** – adapts the instrument to affect the object, *acceptor** – catches the object, *stock** – the object collected for processing, *separator** – sorts the object, *material** – the object used as a raw material for making a product, *model** – the physical or informational original sample for making a product from the object, *retainer** – turns a variable locus of the object into a constant one, *resource** – feeds the instrument, *stimulus** – reveals the parameter of the object, *regulator** – serves as an instruction in making a product from the object, *chronotope** – localizes the object in time, *source** – provides instructions for the instrument, *indicator** – displays a parameter of impact on the object or a parameter of the product as the result of subject's impact on the object.

*Object** – the recipient of the action, varieties of the object: *coating** – the outer insulation of the individ's shell, *hull** – the individ's shell, *interlayer** – the inner insulation of the individ's shell, *kernel** – the core of the individ.

*Product** – the result of the subject's impact (action) on the object (the individ adapted to a given role in a new action), varieties of the product: *billet** – the object turned into a raw material, *semi-product** – the product half-made from raw materials, *prototype** – the prototype product, *end item** – the finished product [61, p. 10, 15-16].

TAPAZ distinguishes between physical and informational processes, since on the highest abstract semantic level the physical action was considered as an influence of one individ onto another through its shell, and the informational action – through its surroundings [44, p. 12], [63], [64]. Below is the classification of semantic elements (macroprocesses) of TAPAZ written by means of the SC-code [44, p. 34], [46, p. 185].

exposure
:= [action]
:= [event]
⊂ *act*
⇒ *explanation**:
[The action is an influence of one individ onto another [44, p. 6].]
⇒ *decomposition**:
{• *activation exposure*
⇒ *decomposition**:
{• *m_perceive*
• *m_reflect*
• *m_comprehend*
• *m_understand*
• *m_attract*
• *m_cumulate*
• *m_constrict*
• *m_attain*
• *m_adopt*
• *m_memorize*
• *m_contemplate*
• *m_learn*
• *m_absorb*
• *m_accumulate*
• *m_center*
• *m_assimilate*
• *m_feel*
• *m_behold*
• *m_feel profoundly*
• *m_experience*
• *m_over absorb*
• *m_concentrate*
• *m_centrifuge*
• *m_dissimilate*
• *m_reject*
• *m_erase*
• *m_rethink*
• *m_overcome*
• *m_expel*
• *m_decompress*
• *m_force off*
• *m_disassociate*
}
• *exploitation exposure*
⇒ *decomposition**:
{• *m_notify*
• *m_advertise*
• *m_instill*
• *m_state*
• *m_approach*
• *m_joint*
• *m_press down*
• *m_connect*
• *m_explain*
• *m_propagandize*
• *m_prove*
• *m_certify*
• *m_insert*
• *m_pump*
• *m_press in*
• *m_link*
• *m_reveal*
• *m_prophesize*
• *m_enlighten*
• *m_divine*
• *m_conduct*
• *m_spread*
• *m_squeeze out*
• *m_disconnect*
• *m_darken*
• *m_encode*
• *m_discredit*
• *m_disavow*
• *m_take out*
• *m_pull up*
• *m_push out*
• *m_unlink*
}
• *transformation exposure*
⇒ *decomposition**:
{• *m_inform*
• *m_interest*
• *m_assure*
• *m_predispose*
• *m_touch on*
• *m_envelope*
• *m_clamp*
• *m_mold*
• *m_admonish*
• *m_teach*
• *m_convince*
• *m_nurture*
• *m_rip up*
• *m_fill up*
• *m_press*
• *m_form*
• *m_pierce*
• *m_intend*
• *m_transfigure*
• *m_reincarnate*
• *m_penetrate*
• *m_overflow*
• *m_unclamp*
• *m_eviscerate*
• *m_pester*
• *m_mesmerize*
• *m_lose conscious*
• *m_go mad*
• *m_punch*

- *m_uplift*
- *m_disband*
- *m_annihilate*

}

- *normalization exposure*

⇒ *decomposition**:

{
- *m_recollect*
- *m_recreate*
- *m_restart*
- *m_render*
- *m_recrystallize*
- *m_reintegrate*
- *m_regenerate*
- *m_restore*
- *m_reproduce*
- *m_reclaim*
- *m_renew*
- *m_revive*
- *m_recuperate*
- *m_rehabilitate*
- *m_reactivate*
- *m_reanimate*

}

}

⇒ *decomposition**:

{
- *surroundings-shell exposure*

⇒ *decomposition**:

{
- *m_perceive*
- *m_reflect*
- *m_comprehend*
- *m_understand*
- *m_attract*
- *m_cumulate*
- *m_constrict*
- *m_attain*
- *m_notify*
- *m_advertise*
- *m_instill*
- *m_state*
- *m_approach*
- *m_joint*
- *m_press down*
- *m_connect*
- *m_inform*
- *m_interest*
- *m_assure*
- *m_predispose*
- *m_touch on*
- *m_envelope*
- *m_clamp*
- *m_mold*
- *m_recollect*
- *m_recreate*
- *m_restart*
- *m_render*
- *m_recrystallize*
- *m_reintegrate*
- *m_regenerate*
- *m_restore*

}

- *shell-core exposure*

⇒ *decomposition**:

{
- *m_adopt*
- *m_memorize*
- *m_contemplate*
- *m_learn*
- *m_absorb*
- *m_accumulate*
- *m_center*
- *m_assimilate*
- *m_explain*
- *m_propagandize*
- *m_prove*
- *m_certify*
- *m_insert*
- *m_pump*
- *m_press in*
- *m_link*
- *m_admonish*
- *m_teach*
- *m_convince*
- *m_nurture*
- *m_rip up*
- *m_fill up*
- *m_press*
- *m_form*
- *m_reproduce*
- *m_reclaim*
- *m_renew*
- *m_revive*
- *m_recuperate*
- *m_rehabilitate*
- *m_reactivate*
- *m_reanimate*

}

- *core-shell exposure*

⇒ *decomposition**:

{
- *m_feel*
- *m_behold*
- *m_feel profoundly*
- *m_experience*
- *m_over absorb*
- *m_concentrate*
- *m_centrifuge*
- *m_dissimilate*
- *m_reveal*
- *m_prophesize*
- *m_enlighten*
- *m_divine*
- *m_conduct*

- *m_spread*
- *m_squeeze out*
- *m_disconnect*
- *m_pierce*
- *m_intend*
- *m_transfigure*
- *m_reincarnate*
- *m_penetrate*
- *m_overflow*
- *m_unclamp*
- *m_eviscerate*

}

- *shell-surroundings exposure*

⇒ *decomposition**:

{
- *m_reject*
- *m_erase*
- *m_rethink*
- *m_overcome*
- *m_expel*
- *m_decompress*
- *m_force off*
- *m_disassociate*
- *m_darken*
- *m_encode*
- *m_discredit*
- *m_disavow*
- *m_take out*
- *m_pull up*
- *m_push out*
- *m_unlink*
- *m_pester*
- *m_mesmerize*
- *m_lose conscious*
- *m_go mad*
- *m_punch*
- *m_uplift*
- *m_disband*
- *m_annihilate*

}

}

⇒ *decomposition**:

{
- *initiation exposure*

⇒ *decomposition**:

{
- *m_perceive*
- *m_attract*
- *m_adopt*
- *m_absorb*
- *m_feel*
- *m_over absorb*
- *m_reject*
- *m_expel*
- *m_notify*
- *m_approach*
- *m_explain*
- *m_insert*
- *m_reveal*
- *m_conduct*
- *m_darken*
- *m_take out*
- *m_inform*
- *m_touch on*
- *m_admonish*
- *m_rip up*
- *m_pierce*
- *m_penetrate*
- *m_pester*
- *m_punch*
- *m_recollect*
- *m_recrystallize*
- *m_reproduce*
- *m_recuperate*

}

- *accumulation exposure*

⇒ *decomposition**:

{
- *m_reflect*
- *m_cumulate*
- *m_memorize*
- *m_accumulate*
- *m_behold*
- *m_concentrate*
- *m_erase*
- *m_decompress*
- *m_advertise*
- *m_joint*
- *m_propagandize*
- *m_pump*
- *m_prophesize*
- *m_spread*
- *m_encode*
- *m_pull up*
- *m_interest*
- *m_envelope*
- *m_teach*
- *m_fill up*
- *m_intend*
- *m_overflow*
- *m_mesmerize*
- *m_uplift*
- *m_recreate*
- *m_reintegrate*
- *m_reclaim*
- *m_rehabilitate*

}

- *amplification exposure*

⇒ *decomposition**:

{
- *m_comprehend*
- *m_constrict*
- *m_contemplate*
- *m_center*
- *m_feel profoundly*

- *m_centrifuge*
- *m_rethink*
- *m_force off*
- *m_instill*
- *m_press down*
- *m_prove*
- *m_press in*
- *m_enlighten*
- *m_squeeze out*
- *m_discredit*
- *m_push out*
- *m_assure*
- *m_clamp*
- *m_convince*
- *m_press*
- *m_transfigure*
- *m_unclamp*
- *m_lose conscious*
- *m_disband*
- *m_restart*
- *m_regenerate*
- *m_renew*
- *m_reactivate*

}

- *generation exposure*
  ⇒ *decomposition**:
  { • *m_understand*
  - *m_attain*
  - *m_learn*
  - *m_assimilate*
  - *m_experience*
  - *m_dissimilate*
  - *m_overcome*
  - *m_disassociate*
  - *m_state*
  - *m_connect*
  - *m_certify*
  - *m_link*
  - *m_divine*
  - *m_disconnect*
  - *m_disavow*
  - *m_unlink*
  - *m_predispose*
  - *m_mold*
  - *m_nurture*
  - *m_form*
  - *m_reincarnate*
  - *m_eviscerate*
  - *m_go mad*
  - *m_annihilate*
  - *m_render*
  - *m_restore*
  - *m_revive*
  - *m_reanimate*

  }

}

⇒ *decomposition**:
{ • *physical action*
  ⇒ *explanation**:
  [The *physical action* is an influence in which the subject's shell acts as an instrument.]
  ⇒ *decomposition**:
  { • *m_attract*
  - *m_cumulate*
  - *m_constrict*
  - *m_attain*
  - *m_absorb*
  - *m_accumulate*
  - *m_center*
  - *m_assimilate*
  - *m_over absorb*
  - *m_concentrate*
  - *m_centrifuge*
  - *m_dissimilate*
  - *m_expel*
  - *m_decompress*
  - *m_force off*
  - *m_disassociate*
  - *m_approach*
  - *m_joint*
  - *m_press down*
  - *m_connect*
  - *m_insert*
  - *m_pump*
  - *m_press in*
  - *m_link*
  - *m_conduct*
  - *m_spread*
  - *m_squeeze out*
  - *m_disconnect*
  - *m_take out*
  - *m_pull up*
  - *m_push out*
  - *m_unlink*
  - *m_touch on*
  - *m_envelope*
  - *m_clamp*
  - *m_mold*
  - *m_rip up*
  - *m_fill up*
  - *m_press*
  - *m_form*
  - *m_penetrate*
  - *m_overflow*
  - *m_unclamp*
  - *m_eviscerate*
  - *m_punch*
  - *m_uplift*
  - *m_disband*

- *m_annihilate*
- *m_recrystallize*
- *m_reintegrate*
- *m_regenerate*
- *m_restore*
- *m_recuperate*
- *m_rehabilitate*
- *m_reactivate*
- *m_reanimate*

}

- *informational action*
  ⇒ *explanation**:
  [The *informational action* is an influence in which the subject's surroundings acts as an instrument.]
  ⇒ *decomposition**:
  {
  - *m_perceive*
  - *m_reflect*
  - *m_comprehend*
  - *m_understand*
  - *m_adopt*
  - *m_memorize*
  - *m_contemplate*
  - *m_learn*
  - *m_feel*
  - *m_behold*
  - *m_feel profoundly*
  - *m_experience*
  - *m_reject*
  - *m_erase*
  - *m_rethink*
  - *m_overcome*
  - *m_notify*
  - *m_advertise*
  - *m_instill*
  - *m_state*
  - *m_explain*
  - *m_propagandize*
  - *m_prove*
  - *m_certify*
  - *m_reveal*
  - *m_prophesize*
  - *m_enlighten*
  - *m_divine*
  - *m_darken*
  - *m_encode*
  - *m_discredit*
  - *m_disavow*
  - *m_inform*
  - *m_interest*
  - *m_assure*
  - *m_predispose*
  - *m_admonish*
  - *m_teach*
  - *m_convince*
  - *m_nurture*
  - *m_pierce*
  - *m_intend*
  - *m_transfigure*
  - *m_reincarnate*
  - *m_pester*
  - *m_mesmerize*
  - *m_lose conscious*
  - *m_go mad*
  - *m_recollect*
  - *m_recreate*
  - *m_restart*
  - *m_render*
  - *m_reproduce*
  - *m_reclaim*
  - *m_renew*
  - *m_revive*

  }

}

The transition can be carried out in 2 stages:

- the transition from the initial version of the text to the reconstructed one;
- the transition from the reconstructed text to semantics.

The reconstruction of the text occurs through the reconstruction of the missing parts of the sentence based on the World Model or the Linguistic Image of the World and then through normalization of its syntactic structure by rewriting the Parts of the Sentence.

During the work, the following rules for the reconstruction of the text were formulated:

- the grammatical direct object of the initial text is displayed in the grammatical predicate of the reconstructed text (for example, the grammatical direct object *производство* = *a production* of the initial text is mapped to the grammatical predicate of the reconstructed text *производим* = *produces*);
- the grammatical attribute of the initial text is mapped to the grammatical direct object of the reconstructed text (for example, the grammatical attribute *творога* = *cottage cheese's* of the initial text is mapped to the grammatical direct object of the reconstructed text *творог* = *cottage cheese*).

The result of reconstruction of the initial text under consideration: «Некто принимает молоко, затем окисляет молоко, а именно: нормализует молоко до 15-процентной жирности, затем очищает молоко, затем пастеризует молоко, затем охлаждает молоко до определённой температуры, затем вносит закваску в молоко, затем сквашивает молоко, затем режет сгусток, затем подогревает сгусток, затем обрабатывает сгусток, затем отделяет сыворотку, затем охлаждает сгусток и, в итоге, производит

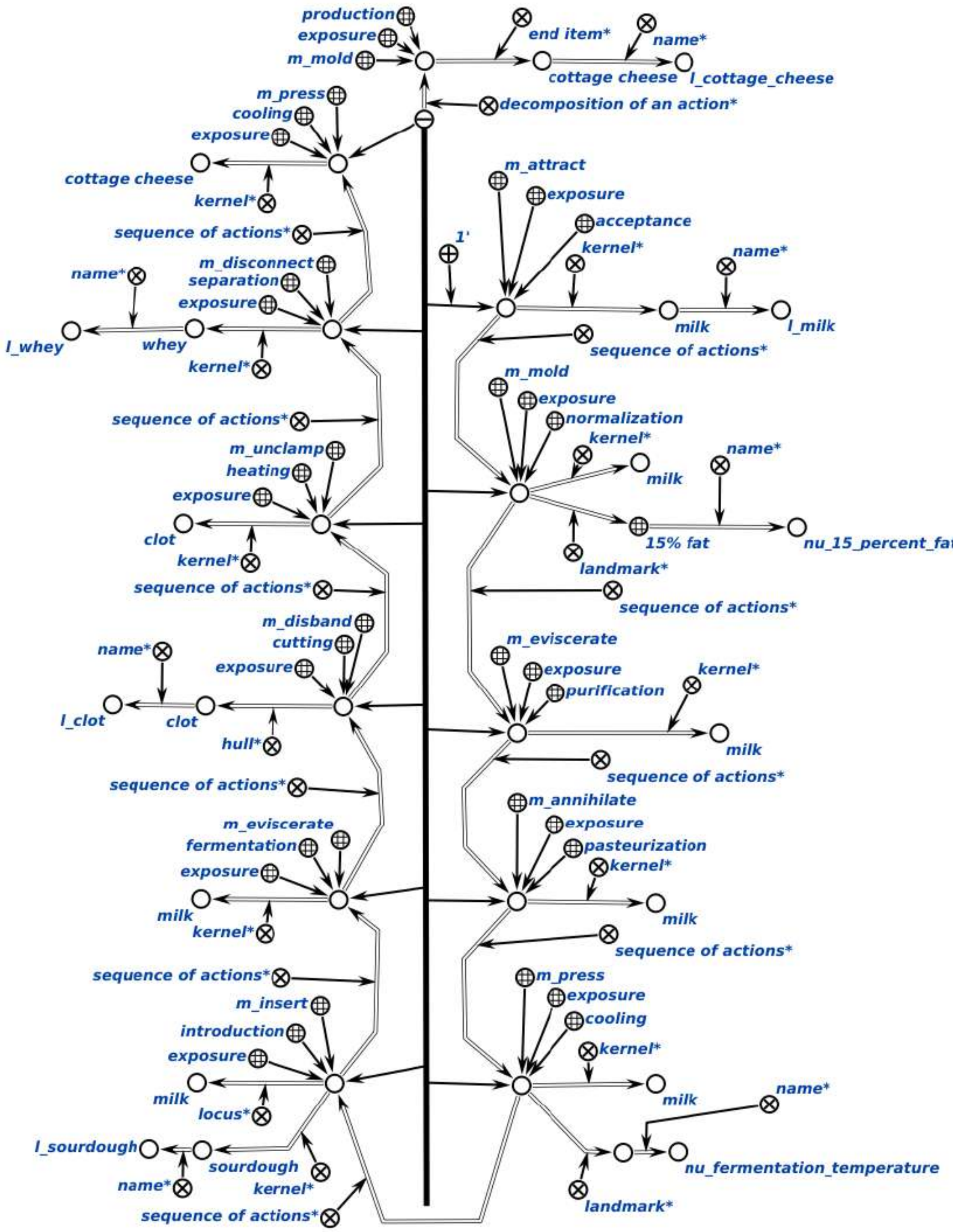

Figure 6. The result of the transition to the terms of TAPAZ-2.

творог»[2].

During the work, the following rules for the transition to semantics were also formulated:

[2] "Someone accepts milk, then acidifies milk, namely: normalizes milk to 15% fat, then purifies milk, then pasteurizes milk, then cools milk to a certain temperature, then adds sourdough to milk, then ferments milk, then cuts the clot, then heats the clot, then processes the clot, then separates whey, then cools the clot and, as a result, produces cottage cheese".

- the grammatical predicate of the reconstructed text is mapped to the action (for example, the grammatical predicate *производит* = *produces*);
- the grammatical direct object of the reconstructed text is mapped to the grammatical object (for example, the grammatical direct object *творог* = *cottage cheese*).

When combining the transition rules from the initial version of the text to the reconstructed one with the rules for the transition to semantics, it is possible to obtain the following rules that provide a one-step transition:

- the grammatical direct object of the initial text is mapped to the action;
- the grammatical attribute of the initial text is mapped to the object.

The final result is shown in Figure 6.

VII. Integration into the knowledge base

The agent of merging structures in the knowledge base integrates the structure obtained as a result of analysis of the text into the knowledge base. The process involves searching for and resolving contradictions.

As an example, we will present a situation when there is a fragment in the knowledge base that describes an exposure that is of the same type as shown in Figure 7.

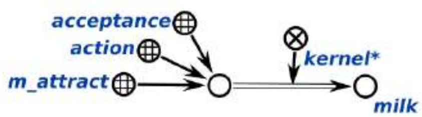

Figure 7. The construction that was present in the knowledge base before merging.

In this case, the models are merged. The resulting construction is shown in Figure 8.

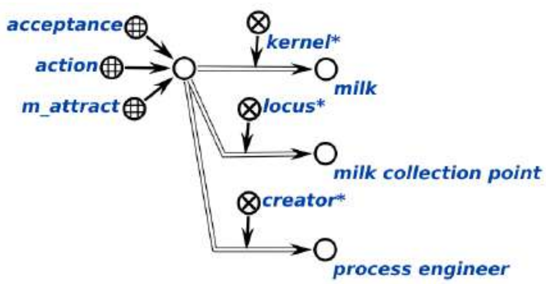

Figure 8. The result of merging constructs.

VIII. Conclusion

The article proposes a new approach to the machine understanding of texts in natural language (Natural Language Understanding, NLU), based on the formalization of the World Model using algorithms of the Theory for Automatic Generation of Knowledge Architecture (TAPAZ-2) and the immersion of the obtained semantic formalisms into the software environment using Open Semantic Technology for Intelligent Systems (OSTIS) that operates with original dynamic graph models – semantic networks in the form of specially oriented taxonomies and ontologies represented in the SC-code (Semantic Computer Code). The resulting taxonomic and ontological set is universal and can be used for machine understanding of collections of texts of various subject domains in various natural languages. The advantages of the approach are:

- decoding the meaning of signs and sense of sentences through decoding the patterns of the World Model, which provides the ability to support analytical activities and solve inventive problems not only by analogy [44, p. 39], [46, p. 192], [61, p. 16];
- standard dynamic graph representation of any type of knowledge within a single knowledge base, regardless of the platform or system [65]–[67];
- a unified top-level algebraic ontology adapted to the semantization of the Internet;
- machine-friendly parsing that provides a straightforward transition to automatic semantic markup of content;
- Semantic Classifier, Role List of Individuals and Knowledge Graph, significantly superior to their analogues in terms of the capacity of semantic calculus [11]–[28], [68], [69];
- mathematical, semantic and software algorithms that can significantly increase the accuracy and speed of operation of search engines;
- compatibility with statistical methods and any types of machine learning that scale the obtained results and reduce the complexity and labor intensity of the knowledge base development.

Acknowledgment

The authors thank the Department of Intelligent Information Technologies of the Belarusian State University of Informatics and Radioelectronics, the Department of Theory and Practice of the Chinese Language of the Minsk State Linguistic University for the help and valuable comments.

References

[1] A. Hardzei. Attributive-Predicative Syntagm as a Language Universal. *Inostrannyye yazyki v vysshey shkole* [Foreign languages in Higher School]. Ryazan, Ryazan State University Publ., 2019, iss. 4(51), pp. 7–19. (In Russian).

[2] A. Hardzei. Foundations of Combinatory Semantics. *Collection of Papers on Lexicography — Slovo i slovar' = Vocabulum et vocabularium* [The Word and the Lexicon]. Grodno, Grodno State University Publ., 2005, pp. 32–35. (In Russian).

[3] A. Hardzei. Linguistic Propaedeutics. *Materialy IV Respublikanskoy nauchnoy konferentsii "Belarus' v sovremennom mire"* [Proceedings of IV Republican Research Conference "Belarus in the Modern World"]. Minsk, The Republican Institute of Higher School Publ., 2005, pp. 226–229. (In Russian).

[4] A. Hardzei. *Printsipy ischisleniya semantiki predmetnykh oblastei* [The Principles of Evaluating the Semantics of Subject Domains]. Minsk, Belarusian State University Publ., 1998, 156 p. (In Russian).

[5] Yu. V. Knorozov. On the Classification of Signaling. *Osnovnyye problemy afrikanistiki* [Basic Problems of African Studies]. Moscow, Science Publ., 1973, pp. 324–334. (In Russian).

[6] E. V. Romanovskaya. The Theory of Fascination in the Context of the Problems of Literary Translation (on the Example of Translations of the Chinese Classical Poetry "Shi"). *Izvestiya Gomel'skogo gosudarstvennogo universiteta imeni F. Skoriny* [Proceedings of Gomel State University named after Francisk Skorina]. Gomel, Gomel State University Publ., 2019, no. 1(112), pp. 125–131. (In Russian).

[7] A. Hardzei. About Declarative and Procedural Knowledge Representation. *Inostrannyye yazyki v vysshey shkole* [Foreign languages in Higher School]. Ryazan, Ryazan State University Publ., 2021, iss. 3(58), in print. (In Russian).

[8] M. Minsky. A Framework for Representing Knowledge. *MIT-AI Laboratory*, memo no 306, June, 1974, pp. 1–81. [Online]. Available at: https://dspace.mit.edu/bitstream/handle/1721.1/6089/AIM306.pdf?sequence_2, (accessed 29.07.2021).

[9] V. V. Martynov. *V tsentre soznaniya cheloveka* [In the Center of Human Conscious]. Minsk, Belarusian State University Publ., 2009, 272 p. Available at: https://dokumen.tips/documents/-55cf9b12550346d033a49d65.html, (accessed 29.07.2021).

[10] Z. S. Harris. Co-occurrence and Transformation in Linguistic Structure. *Language*. Baltimore, 1957, vol. 33, no 3, pp. 283–340.

[11] Ch. J. Fillmore. The Case for Case. *Universals in Linguistic Theory*. New-York, Harper and Row, 1968, pp. 1–88.

[12] Ch. J. Fillmore. Types of Lexical Information. *Studies in Syntax and Semantics*. Dordrecht, Reidel, 1969, pp. 109–137.

[13] Ch. J. Fillmore. Verbs of Judging: An Exercise in Semantic Description. *Studies in Linguistic Semantics*. New York, Holt, Rinehart and Winston, 1971, pp. 272–289.

[14] Ch. J. Fillmore. Verbs of Judging: An Exercise in Semantic Description. *Studies in Linguistic Semantics*. New York, Holt, Rinehart and Winston, 1971, pp. 272–289.

[15] Ch. J. Fillmore. The Case for Case Reopened. *Syntax and Semantics*. New-York, Academic Press, 1977, vol. 8, pp. 59–81.

[16] Ch. J. Fillmore. Frame Semantics. *Linguistics in the Morning Calm*. Papers Prepared for the 1981 Seoul International Conference on Linguistics. Seoul, Hanshin Publishing Co., 1982, pp. 111–137.

[17] Ch. J. Fillmore. Frames and the Semantics of Understanding. *Quaderni di Semantica*. 1985, vol. 6, no. 2, pp. 222–254.

[18] Ch. J. Fillmore. "Corpus Linguistics" vs. "Computer-aided Armchair Linguistics". *Directions in Corpus Linguistics*. Proceedings from a 1991 Nobel Symposium on Corpus Linguistics. Stockholm, Mouton de Gruyter, 1992, pp. 35–66.

[19] Ch. J. Fillmore. Encounters with Language. *Computational Linguistics*. 2012, vol. 38 (4), pp. 701–718.

[20] Ch. J. Fillmore, B. T. S. Atkins. Towards a Frame-based Lexicon: The Semantics of RISK and Its Neighbors. *Frames, Fields and Contrasts*. Hillsdale (NJ), Lawrence Erlbaum Associates, 1992, pp. 75–102.

[21] Ch. J. Fillmore, B. T. S. Atkins. Starting where the Dictionaries Stop: The Challenge for Computational Lexicography. *Computational Approaches to the Lexicon*. Oxford, Oxford Univ. Press, 1994, pp. 349–393.

[22] Ch. J. Fillmore, B. T. S. Atkins. *Polysemy: Theoretical and Computational Approaches*. Oxford, Oxford Univ. Press, 2000, pp. 91–110.

[23] T. A. Stowell. *Origins of phrase structure*. Thesis (Ph. D.) Massachusetts Institute of Technology, Dept. of Linguistics and Philosophy, 1981, 496 p. Available at: http://www.ai.mit.edu/projects/dm/theses/stowell81.pdf, (accessed 29.07.2021).

[24] T. Berners-Lee, J. Hendler, O. Lassila. The Semantic Web: A New Form of Web Content that is Meaningful to Computers Will Unleash a Revolution of New Possibilities. *Scientific American*. 2001, vol. 284, iss. 5, pp. 34–43.

[25] T. Berners-Lee, N. Shedbolt, W. Hall. Semantic Web Revisited. *IEEE Intelligent Systems*. 2006, vol. 21, iss. 3, pp. 96–101.

[26] Activity Vocabulary. W3C Recommendation. Available at: https://www.w3.org/TR/activitystreamsvocabulary/#dfn-activity, (accessed 29.07.2021).

[27] Semantic Web. Available at: https://www.w3.org/standards/semanticweb, (accessed 29.07.2021).

[28] Schema.org Vocabulary, version 8.0. Available at: https://schema.org, (accessed 29.07.2021).

[29] S. P. Rippa, O. M. Lyashenko. Semantic platforms of knowledge bases software in informatics. *Pratsi Odes'koho politekhnichnoho universytetu* [Proceedings of Odessa Polytechnic University]. Odessa, Odessa Polytechnic University Publ., 2013, vol. 1(40), pp. 91–96. (In Russian).

[30] V. V. Martynov. *Osnovy semanticheskogo kodirovaniya. Opyt predstavleniya i preobrazovaniya znaniy* [Foundations of Semantic Coding. Experience of Knowledge Representation and Transformation]. Minsk, European Humanitarian University Publ., 2001, 140 p. (In Russian).

[31] D. G. Bobrow, T. Winograd. An Overview of KRL, a Knowledge Representation Language. *Cognitive Science*. 1977, vol. 1 (1), pp. 3–46.

[32] A. Vežbicka. From the Book "Semantic Primitives". *Semiotika. Antologiya* [Semiotics. Anthology] / comp. by Yu. S. Stepanov. Moscow, Academichesky Proekt; Yekaterinburg, Delovaya Kniga, 2001, pp 42–270. (In Russian).

[33] Ashtadhyayi or Sutrapath of Panini. Available at: https://sanskritdocuments.org/doc_z_misc_major_works/aShTAdhyAyI.pdf, (accessed 29.07.2021).

[34] V. N. Toporov. On some analogies to the problems and methods of modern linguistics in the works of ancient Indian grammarians. *Kratkiye soobshcheniya Instituta narodov Azii* [Briefs of the Institute of the Peoples of Asia]. Moscow, USSR Academy of Sciences Publ., 1961, iss. 57, pp. 123–133. (In Russian).

[35] O. A. Voloshina. On the structure and linguistic terminology of Panini's grammar and its influence on European linguistics. *Vestnik RGGU. Seriya: Istoriya. Filologiya. Kul'turologiya. Vostokovedeniye* [Bulletin of the Russian State University for the Humanities. Series: History. Philology. Culturology. Oriental studies]. 2010, no. 9 (52), pp. 161–172. (In Russian).

[36] V. V. Martynov. *Semiologicheskie osnovy informatiki* [Semeiological Foundations of Computer Science]. Minsk, Science and Technology Publ., 1974, 192 p. (In Russian).

[37] V. V. Martynov. *Universal'nyy semanticheskiy kod. Grammatika. Slovar'. Teksty* [Universal Semantic Code. Grammar. Dictionary, Texts]. Minsk, Science and Technics Publ., 1977, 191 p. (In Russian).

[38] V. V. Martynov. *Universal'nyy semanticheskiy kod: USK-3* [Universal Semantic Code: USC–3]. Minsk, Science and Technics Publ., 1984, 134 p. (In Russian).

[39] V. V. Martynov. *Universal'nyy semanticheskiy kod: USK–4* [Universal Semantic Code: USC–4]. Preprint, no 2 / Institute of Linguistics of Academy of Science of BSSR. Minsk, Science and Technics Publ., 1988, 30 p. (In Russian).

[40] V. V. Martynov. *Universal Semantic Code: USC–5*. Minsk, Minsk State Linguistic University, 1995, 21 p. (Pre-print. MSLU; no. 4).

[41] A. Hardzei. Procedural Semantics and Evaluation of Subject Domains. *Materialy I Mezhdunarodnoy nauchnoy konferentsii "Yazyk: semantika, sintaktika, pragmatika"* [Proceedings of I International Research Conference "Language: semantics, syntactics, pragmatics". Minsk, Minsk State Linguistic University Publ., 1995, pt. 1, pp. 16–17. (In Russian).

[42] A. Hardzei. *Deduktivnaya teoriya yazyka* [Deductive Theory of Language]. Minsk, Belarussian Science Publ., 1998, 46 p. (In Russian).

[43] A. Hardzei. Theory for Automatic Generation of Knowledge Architecture (TAPAZ-2) and Further Minimization of Semantic Calculus. *Materialy IV Mezhdunarodnoy nauchno-tekhnicheskoy konferentsii "Otkrytyye semanticheskiye tekhnologii proyektirovaniya intellektual'ykh system (OSTIS-2014)"* [Proceedings of IV International Scientific and Technical Conference "Open Semantic Technologies for Intelligent Systems (OSTIS-2014)"]. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2014, pp. 49–64.

[44] A. Hardzei. *Theory for Automatic Generation of Knowledge Architecture: TAPAZ-2* / Transl. from Rus. I. M. Boyko. Rev. English edn. Minsk, The Republican Institute of Higher School Publ., 2017, 50 p. Available at: http://tapaz.by, (accessed 18.04.2021).

[45] A. Hardzei, A. Udovichenko. Graph of TAPAZ–2 Semantic Classifier. *Open Semantic Technologies for Intelligent Systems (OSTIS)*. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2019, iss. 3, pp. 281–284.

[46] A. Hardzei. Plagiarism Problem Solving Based on Combinatory Semantics. *Communications in Computer and Information Science (CCIS)*. Switzerland, Springer Nature Switzerland AG, 2020, vol. 1282, pp. 176–197. Available at: https://link.springer.com/book/10.1007%2F978-3-030-60447-9), (accessed 29.07.2021).

[47] V. Golenkov, N. Gulyakina, N. Grakova, I. Davydenko, V. Nikulenka, A. Eremeev, V. Tarasov. From Training Intelligent Systems to Training Their Development Tools. *Open Semantic Technologies*

for Intelligent Systems (OSTIS). Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2018, iss. 2, pp. 81–98.
[48] T. N. Domnina, O. A. Khachko. *Nauchnyye zhurnaly: kolichestvo, tempy rosta* [Scientific Journals: Quantity, Growth Rates]. Available at: http://www.benran.ru/SEM/Sb_15/sbornik/ 83.pdf, (accessed: 29.07.2021). (In Russian).
[49] T. V. Chernigovskaya. *Kak internet izmenil nash mozg* [How the Internet Has Changed Our Brain]. Available at: https://golbis.com/pin/neyrolingvist-tatyana-chernigovskaya-kakinternet-vliyaet-na-nash-mozg-8, (accessed: 29.07.2021). (In Russian).
[50] S. Ivanov. *Kolichestvo publikatsiy nauchnykh rabot v mire yezhegodno uvelichivayetsya, prichem osnovnoy protsent rosta prikhoditsya imenno na razvivayushchiyesya strany* [The number of Scientific Publications in the World Is Increasing Every Year, with the Bulk of the Growth Accounted for by Developing Countries]. Available at: https://hightech.fm/2018/12/24/science-2, (accessed: 29.07.2021). (In Russian).
[51] V. Golenkov, N. Guliakina, I. Davydenko, A. Eremeev. Methods and tools for ensuring compatibility of computer systems. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem* [Open Semantic Technologies for Intelligent Systems (OSTIS)]. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2019, iss. 3, pp. 25–52.
[52] A. Hardzei. Parts of Language Instead of Parts of Speech. *Yazyk. Glagol. Predlozheniye* [Language. Verb. Sentence]. Smolensk, Smolensk Pedagogical State University Publ., 2000, pp. 258–271. (In Russian).
[53] A. Hardzei. The Paradigm of Parts of the Language. *Materially VIII Mezhdunarodnoi konferentsii —Slovoobrazovaniye i nominativenaya derivatsiya v slavyanskikh yazykakh* [International Research Conference —Word-Building and Nominative Derivation in Slavic Languages]. Grodno, Grodno State University Publ., 2003, pp. 173–179. (In Russian).
[54] A. Hardzei. Parts of Language and the Procedures of Its Delineation. *Puti Podnebesnoi* [The Paths of the Middle Kingdom]. Minsk, Belarusian State University Publ., 2006, iss. 1, pt. 1, pp. 69–75. (In Russian).
[55] A. Hardzei. The Meaning of Logical and Semantic Paradoxes for Language Theory. *Karpovskiye nauchnye chteniya* [Readings from V. A. Karpov]. Collection of Research Papers in 2 parts. Minsk, Belarussian State University Publ., 2011, pt. 1, pp. 9–18. (In Russian).
[56] A. Hardzei, Sun Kewen. Typological Aspects of Word Formation. *Puti Podnebesnoi* [The Paths of the Middle Kingdom]. Minsk, The Republican Institute of Higher School Publ., 2011, iss. 2, pp. 40–53. (In Russian).
[57] A. Hardzei. Virtual String as a Syntactical Code of a Sentence (by the Examples of the Chinese Language). *Materialy mezhdunarodnoy nauchnoy konferentsii "Yazyk, obshchestvo i problem mezhkul'turnoy kommunikatsii"* [Proceedings of International Research Conference "Language, Society and Problems of Intercultural Communication"]. Grodno, Grodno State University Publ., 2007, pt. 2, pp. 349–358. (In Russian).
[58] A. Hardzei. About the Combinatorics of Numerical Signs in the Chinese Language. *Nauchnyye chteniya, posvyashchennyye Viktoru Vladimirovichu Martynovu* [Scientific readings dedicated to Viktor Vladimirovich Martynov]. Collection of Research Papers. Minsk, The Republican Institute of Higher School Publ., 2015, iss. 2, pp. 12–14.
[59] A. Hardzei. Dynamic Syntax: A Semantic View. *Inostrannyye yazyki v vysshey shkole* [Foreign languages in Higher School]. Ryazan, Ryazan State University Publ., 2017, iss. 4 (43), pp. 26–34. (In Russian).
[60] A. Hardzei. *Teoreticheskaya grammatika vostochnykh yazykov: lektsionnyy kurs* [Theoretical Grammar of Oriental Languages: Lecture Course]. Electronic data. Minsk, 2007, 1 electronic optical disc (CD-ROM).
[61] A. Hardzei. Semantic Markup of the Event and its Display by Means of the Chinese and Russian Languages. *Inostrannyye yazyki v vysshey shkole* [Foreign languages in Higher School]. Ryazan, Ryazan State University Publ., 2021, iss. 2 (57), pp. 5–26. (In Russian).
[62] M. I. Svyatoshchik. Translating Multi-Object Semantics into SingleObject Semantics as a Method of Minimizing Semantic Calculation. *Uchonyye zapiski VGU im. P. M. Masherova: sb. nauch. trudov [Scientific notes of Vitebsk State University named after P. M. Masherov*: collection of scientific papers]. Vitebsk, *Vitebsk State University Publ.*, 2020, vol. 31, pp. 180–184. (In Russian).
[63] A. Hardzei. Cognitive Approach to Teaching of Signs. *Komp'yuternyye programmy v obuchenii belorusskomu i inostrannym yazykam: Tezisy dokladov nauchauchnoy konferentsii* [Computer Programs in Teaching Byelorussian and Foreign Languages: Abstracts of the Research Conference]. Minsk, Minsk State Linguistic University Publ., 1995, pp. 18–20. (In Russian).
[64] A. Hardzei. Procedural Semantics and Calculus of Subject Domains. *Yazyk: semantika, sintaktika, pragmatika: Tezisy dokladov nauchnoy konferentsii* [Language: Semantics, Syntactics, Pragmatics: Abstracts of the Research Conference]. Minsk, Minsk State Linguistic University Publ., 1995, pt. 1, pp. 16–17. (In Russian).
[65] I. T. Davydenko. Semantic Models, Method and Tools of Knowledge Bases Coordinated Development Based on Reusable Components. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem* [Open Semantic Technologies for Intelligent Systems (OSTIS-2018)]. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2018, iss. 2, pp. 99–118.
[66] D. Shunkevich. Agent-oriented Models, Method and Tools of Compatible Problem Solvers Development for Intelligent Systems. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem* [Open Semantic Technologies for Intelligent Systems (OSTIS-2018)]. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2018, iss. 2, pp. 119–132.
[67] I. T. Davydenko. Ontology-Based Knowledge Base Design. *Otkrytye semanticheskie tekhnologii proektirovaniya intellektual'nykh sistem* [Open Semantic Technologies for Intelligent Systems (OSTIS-2017)]. Minsk, Belarussian State University of Informatics and Radioelectronics Publ., 2017, iss. 1, pp. 57–72.
[68] R. Jackendoff. *Semantic Interpretation in Generative Grammar*. Cambridge, MA, MIT Press., 1972, 400 p.
[69] R. Jackendoff. *X–bar Syntax: A Study of Phrase Structure*. Cambridge, MA, MIT Press., 1977, 249 p.
[70] V. V. Golenkov, N. A. Gulyakina. A Project of an Open Semantic Technology for the Component Design of Intelligent Systems. Part 1: Principles of Creation. *Ontologiya proyektirovaniya* [Ontology of Design]. Samara, Samara National Research University Publ., 2014, no. 1, pp. 42–64. (In Russian).
[71] V. V. Golenkov, N. A. Gulyakina. A Project of an Open Semantic Technology for the Component Design of Intelligent Systems. Part 2: Unified Design Models. *Ontologiya proyektirovaniya* [Ontology of Design]. Samara, Samara National Research University Publ., 2014, no. 4, pp. 34–53. (In Russian).

Обработка и понимание естественного языка интеллектуальной системой

Гордей А. Н., Святощик М. И.,
Бобёр Е. С., Никифоров С. А.

Статья посвящена обработке данных на естественном языке в парадигме Теории автоматического порождения архитектуры знаний (TAPAZ-2) и погружении полученных семантических формализмов в программную среду посредством Открытой семантической технологии для интеллектуальных систем (OSTIS). Особенностью подхода является формализация семантики естественного языка с опорой на модель мира и сочетание семантического кодирования с онтологией и таксономией семантических сетей.

Received 01.06.2021