Chapter 2
Introduction to Integral Formal Semantics of Natural Language

Abstract This chapter sets forth the basic ideas and components of Integral Formal Semantics (IFS) of Natural Language – a many-component branch both of formal semantics of NL and Computer Science developed by the author of this book. Section 2.1 describes the basic principles of IFS and introduces the notion of a broadly applicable conceptual metagrammar. Section 2.2 shortly characterizes the principal components of IFS. Sections 2.3, 2.4, 2.5, 2.6, and 2.7 describe a number of the principal components of IFS. These sections contain numerous examples reflecting the different stages of elaborating powerful and flexible formal means for describing semantic structure of NL-texts – sentences and discourses.

2.1 The Basic Principles of Integral Formal Semantics of Natural Language

Integral Formal Semantics of Natural Language (IFS) is a many-component branch both of formal semantics of NL and of the theory of natural language processing systems as a part of Computer Science. It consists of several theories, mathematical models, and algorithms developed by the author of this monograph since the beginning of the 1980s.

2.1.1 Basic Principles

The basic principles of IFS stated below correspond very well to the requirements of Cognitive Linguistics and Computer Science concerning the formal study of the regularities of conveying information by means of NL. IFS proposes, first of all, a new class of formal systems for building semantic representations of sentences and discourses with high expressive power being close to the expressive power of NL.

The total content of the next chapters of this book can be considered as the kernel of the current configuration of IFS.

The basic principles of IFS are as follows:

1. The main goal of the researches on the formalization of NL-semantics is to be the construction of formal models of Natural Language Processing Systems (NLPSs) and of such subsystems of NLPSs which belong to the so-called semantic components of NLPSs. This means that the accent in the researches is to be on modeling the regularities of the communication of intelligent systems by means of NL.

2. The studies are to be oriented toward considering not only the assertions but also the commands, questions, and discourses which may be the inputs of NLPSs.

3. The basis of the studies is to be a formal model reflecting many peculiarities of semantic structures of sentences and discourses of arbitrary big length and providing a description of some class *Langsem* of formal languages being convenient for building semantic representations (SRs) of NL-texts in a broad spectrum of applications and on different levels of representation.

4. The central roles in the development of formal models for the design of NLPSs must play the models of the following correspondences:

   - "NL-text or its special representation (e.g., a marked-up representation of a text) + Knowledge ⇔ Semantic representation of a text" (both for the analyzers and generators of NL);
   - "An NL-text or its special (marked-up) representation + Knowledge → Semantic representation of a text + Plan of the reaction" for designing NL-interfaces to the recommender systems, expert systems, personal robots, etc. (the reactions may be questions, movements, calculations, etc.);
   - "Text of a request or its special (marked-up) representation + Text of an information source (or a semantic representation of the latter text) + Knowledge → A textual or semantic representation of retrieved information or Negative answer" for designing full-text databases and the systems which automatically form and update the knowledge bases of applied intelligent systems.

5. The model-theoretical semantics of NL is to play the auxiliary roles. The first to third sections of this chapter and the papers [58, 64, 65] contain the proposals concerning the formal structure of models of the listed kinds.

6. Semantics and pragmatics of NL should be studied jointly by means of the same formal techniques. It should be noted that this principle underlies the works [52, 53, 55, 56]. Hence this principle was formulated several years before the publication of the works [180, 189, 190].

7. A formal description of the surface structure of any NL-text T is to be based on a formal description of the structured meaning of T and on a formal description of the semantic – syntactic structure of T. Purely syntactic descriptions of texts' structures may be useful, but are not necessary. Such syntactic descriptions are to be the derivatives of the descriptions of semantic and semantic–syntactic

structures of NL-texts. This point of view is directly opposite to the approach used, in particular, in Montague Grammar and in Generalized Phrase Structure Grammars. However, it seems that the suggested viewpoint is (a) more similar to the processes realizing in the course of human thinking; (b) more practically effective; and (c) the only useful as concerns describing surface structure of scientific articles, books, etc. The stated principle may be considered as a possible formulation of one of the key ideas of Cognitive Linguistics. This idea set forth, in particular, in [20, 149, 152, 187] is the dependency of syntax on semantics.

8. The semantic interpretation of a phrase being a fragment of a published discourse is to depend on the knowledge about reality, on the source where the discourse is published, and on the meanings of precedent fragments of the discourse (in some cases – on the meanings of some next fragments too).

9. The semantic interpretation of an utterance in the course of a dialogue is to depend in general case on the knowledge about reality, about dialogue participants (in particular, about their goals), about the discussed situation, and about the meanings of previous utterances.

10. The languages from the class *Langsem* are to provide the possibility to represent knowledge about the reality and, in particular, to build formal descriptions of notions and regularities and also the descriptions of the goals of intelligent systems and of the destinations of things.

11. The languages from the class *Langsem* are to give the opportunity to represent the knowledge modules (blocks, chunks in other terms) as the units having some external characteristics (Authors, Date, Application domains, etc.) or metadata.

12. The languages from the class *Langsem* are to allow for building the models of structured hierarchical conceptual memory of applied intelligent systems, the frame-like representations of knowledge and are to be convenient for describing the interrelations of knowledge modules.

2.1.2 The Notion of a Broadly Applicable Conceptual Metagrammar

Let's call a formal model of the kind described above in the principle 3 a Broadly Applicable Metagrammar of Conceptual Structures or a Broadly Applicable Conceptual Metagrammar (BACM).

A Broadly Applicable Conceptual Metagrammar should enable us to build formal semantic analogues of sentences and discourses; hence the expressive power of formal languages determined by the model may be very close to the expressive power of NL (if we take into account the surface semantic structure of NL- texts). Besides, a BACM is to be convenient for describing various knowledge about the world [52, 54–56, 62, 63, 65, 67, 68].

If a model is convenient for describing arbitrary conceptual structures of NL-texts and for representing arbitrary knowledge about the world, we say about a Universal Metagrammar of Conceptual Structures or a Universal Conceptual Metagrammar (UCM) .

The reason to say about a metagrammar but not about a grammar is as follows: A grammar of conceptual structures is to be a formal model dealing with the elements directly corresponding to some basic conceptual items (like "physical object," "space location")

An example of such semi-formal grammar is provided by the known Conceptual Dependency theory of Schank. On the contrary, a metagrammar of conceptual structures is to postulate the existence of some classes of conceptual items, to associate in a formal way with arbitrary element from each class certain specific information, and to describe the rules to construct arbitrarily complicated structured conceptual items in a number of steps in accordance with such rules (proceeding from elementary conceptual items and specific information associated with arbitrary elements of considered classes of items).

The most part of the known approaches to the formalization of NL-semantics practically doesn't give the cues for the construction of an UCM. This applies, in particular, to Montague Grammar, Discourse Representation Theory (DRT), Theory of Generalized Quantifiers, Situation Theory, Dynamic Montague Grammar, Dynamic Predicate Logic, Theory of Conceptual Graphs, and Episodic Logic.

For instance, it is difficult to not agree with the opinion of Ahrenberg that "in spite of its name, DRT can basically be described as formal semantics for short sentence sequences rather than as a theory of discourse" [3]. This opinion seems to be true also with respect to the content of the monograph [143].

Happily, a considerable contribution to outlining the contours of a Universal Conceptual Metagrammar has been made by Integral Formal Semantics of NL.

2.2 The Components of Integral Formal Semantics of Natural Language

In order to list the principal components of IFS, we need the notion of a formal system, or a calculus. In discrete mathematics, the development and investigation of formal systems, or calculuses, is the main manner of studying the structure of strings belonging to formal languages.

Following [194], by *a formal system*, or *a calculus*, we'll mean any ordered triple

$$F = (L, L_0, R),$$

where L is a formal language in an alphabet, $L_0 \subset L$, R is a finite set of rules enabling us to obtain from the strings of L another strings of L. The rules from R are called *the inference rules.*

The strings of L that one can obtain as a result of applying the rules from R and starting with the strings from L_0 are called the formulas of the system F. We will be interested in what follows in such an interpretation of a calculus when formulas are considered not as theorems but as expressions of some language (in applications – as semantic representations (SRs) of texts and as parts of SRs).

It should be mentioned in this connection that for every context-free grammar generating a language L, one can easily define such calculus that the set of its formulas will be L.

The principal components of IFS are as follows:

1. The theory of S-calculuses and S-languages (the SCL-theory) developed in the first half of 1982 and proposed the new formal means for describing both separate sentences and complex discourses in NL of arbitrary big length (see Sect. 2.3).

2. A mathematical model of a correspondence between the NL-texts (sentences and discourses expressing the commands to a dynamic intelligent device or the commands to draw the geometric figures) and their semantic representations being the strings of restricted S-languages (see Sect. 2.4).

3. The theory of T-calculuses and T-languages (the TCL-theory) studying the semantic structure of discourses introducing a new notion or a new designation of an object (see Sect. 2.5).

4. The initial version of the theory of K-calculuses (knowledge calculuses) and K-languages (knowledge languages), or the KCL-theory, is a new step (in comparision with the SCL-theory) on the way of creating the formal means convenient for describing semantic structure of both sentences and complex discourses in NL (see Sect. 2.6).

5. The current version of the theory of K-calculuses and K-languages (its kernel is the theory of SK-languages – standard knowledge languages) set forth in [85, 91] and in Chaps. 2, 3, 4 and 5 of this book.

6. The analysis of the possibilities to use the theory of SK-languages for solving a number of significant problems of modern Computer Science and Web Science (see Chap. 6 of this monograph).

7. A broadly applicable mathematical model of a linguistic database, that is, a model of a collection of semantic-syntactic data associated with primary lexical units and used by the algorithms of semantic-syntactic analysis for building semantic representations of natural language texts (see Chap. 7 of this book).

8. A new method of transforming an NL-text (a statement, a command, or a question) into its semantic representation (see Chap. 8 of this book).

9. Two complex, strongly structured algorithms of semantic-syntactic analysis of NL-texts (they possess numerous common features). The first one is described in the book [85] and the second one is proposed in Chaps. 9 and 10 of this monograph.

10. The proposals concerning the structure of formal models being useful for the design of semantics-oriented NLPSs (Chap. 1 of this book and [64, 65]).

The components 5–10 of Integral Formal Semantics of Natural Language form the theory of K-representations (knowledge representations). The principal part of the theory of K-representations is set forth in this monograph.

2.3 The Theory of S-Calculuses and S-Languages

The theory of S-calculuses and S-languages (the SCL-theory) is set forth in the publications [52, 53, 55, 56] and in the Ph.D. dissertation [54]. This theory proposed already in 1981–1983 is a really ecological approach to the formalization of NL-semantics, providing powerful and convenient mathematical means for representing both structured meanings of NL-texts and knowledge about the reality.

The basic ideas of the SCL-theory were presented, in particular, at the First symposium of the International Federation of Automatic Control (IFAC) on Artificial Intelligence, which was held in 1983 in Sankt-Petersburg, Russia, and were published in the proceedings of this symposium; it should be noted that the paper [56], published by Pergamon Press, is a considerably abridged version of the publication [55].

The principal part of the SCL-theory is a new formal approach (new for the beginning of the 1980s) to describing conceptual (or semantic) structure of sentences and discourses in NL. The paper [52] for the first time in the world stated the task of developing mathematical models destined for describing structured meanings not only of sentences but also of complicated discourses in NL. Besides, this paper proposed the schemas of 16 partial operations on the finite sequences consisting of conceptual structures associated with NL-texts.

Example 1. Let T1 be the discourse "Sergey and Andrey are friends of Igor and are the physicists. He had told them that he didn't want to work as a programmer. Sergey believed that it would be useful for Igor to have a talk with the Associate Professor Somov and advised him to act in such a way."

In the paper [52], it was proposed to associate the discourse T1 with the following semantic representation $Semrepr1$:

$$((((((\{\downarrow man * Name(\Delta 1, Sergey) \ : \ x1,$$

$$\downarrow man * Name(\Delta 1, Andrey) \ : \ x2\} \ = \ M1) \wedge$$

$$Subset(M1, Friends(\downarrow man * Name(\Delta 1, Igor) \ : \ x3))) \wedge$$

$$(Profession((x1 \wedge x2)) \ = \ physicist)) \wedge$$

$$((P1 \ = \ \neg Want[time-gramm, past](Subject, x3)(Goal,$$

$$Work(Qualification, programmer)(Institution,$$

$$\downarrow res-inst * Name((\Delta 1, Plastics Research Institute))))) \wedge$$

$$Say[time-gramm, past](Subject, x3)(Addressees, M1)$$

$$(Proposition,\ P1)\ ::\ e1)) \wedge ((Believe[time-gramm,\ past](Subject,\ x1)$$

$$(Proposition,\ Useful[time-gramm,\ future](Person1,\ x3)$$

$$(Goal,\ Talk(Person2, \downarrow man*(Title((\Delta 1,\ Assoc-Professor)$$

$$\wedge Surname((\Delta 1,\ Somov)))\ :\ G1))\ ::\ e2 \wedge Cause(el,\ e2)) \wedge$$

$$(Advise[time-gramm,\ past](Subject,\ x1)(Addressees,\ x3)$$

$$(Goal,\ G1)\ ::\ e3 \wedge Cause(e2,\ e3)))).$$

Let's pay attention to the peculiarities of this formal expression being new at the time of publishing the paper [52]. We can find in this formal expression the following original features:

- the compound designations of the notions $man*Name(\Delta 1,\ Sergey)$, $man*Name(\Delta 1,\ Andrey)$, $man*Name(\Delta 1,\ Igor)$;
- the compound designations of the concrete persons with the names *Sergey, Andrey, Igor*;
- a description of a set

$$(\{\downarrow man*Name(\Delta 1,\ Sergey)\ :\ x1,$$

$$\downarrow man*Name(\Delta 1,\ Andrey)\ :\ x2\}\ =\ M1);$$

- a formal representation of the meaning of a sentence with indirect speech;
- the substring $(x1 \wedge x2)$, where the logical connective $\wedge$ (conjunction, and) joins the designations of the persons $X1$ and $X2$ (but not the formulas representing propositions as in first-order logic);
- the substrings of the forms $Expression1 :: e1$, $Expression2 :: e2$, and $Expression3 :: e3$, where $Expression1$, $Expression2$, and $Expression3$ are the descriptions of some events, and $e1$, $e2$, $e3$ are the marks of these events;
- a compound designation of a goal to have a talk with the Associate Professor Somov

$$Talk(Person2, \downarrow man*(Title(\Delta 1,\ Assoc-Professor)$$

$$\wedge Surname(\Delta 1,\ Somov)))\ :\ G1;$$

- the compact representations of causal relationships $Cause(e1, e2)$ and $Cause(e2, e3)$, constructed due to the association of the marks $e1$, $e2$, $e3$ with the descriptions of concrete events in the left fragments of the semantic representation $Semrepr1$.

Example 2. Let T2 be the question "What did Igor say, and to whom did he tell it ?", Then, according to [52], the formula

$$?Transfer-information[time,\ past](Subject, \downarrow man*Name(\Delta 1,\ Igor))$$

$$(Mode1,\ voice)(Addressees,\ ?y1)(Proposition,\ ?p1)$$

may be regarded as a possible semantic representation of T2.

The ideas of the papers [52, 53] received a mathematical embodiment in the Ph.D. dissertation [54]. This dissertation contains a complete mathematical model of a system consisting of 14 partial operations on the finite sequences consisting of conceptual structures associated with NL-texts.

Example 3. Let *Seq*1 be the sequence consisting of the informational units $\vee$, *airplane*, *helicopter*, *dirigible*, *glider*, *deltaplane*. Then one of these partial operations allows for constructing the formal expression

$$(airplane \vee helicopter \vee dirigible \vee glider \vee deltaplane),$$

considered as the value of this operation on the sequence *Seq*1.

The mathematical model constructed in the Ph.D. dissertation [54] defines the formal systems (or calculuses) of four new kinds (the S-calculuses of types 1–4) and, as a consequence, the formal languages of four new kinds (the restricted S-languages of types 1–4) . The S-calculuses of types 1–3 and the restricted S-languages of types 1–3 were determined as preliminary results in order to achieve the final goal: the definition of the class of restricted S-languages of type 4.

Some denotations introduced in [54] are different from the denotations used in [52]. In particular, the expressions of the form $\{d_1, \ldots, d_n\}$, used in [52] for denoting the sets consisting of the objects $d_1, \ldots, d_n$, are not employed in [54].

For instance, the expression

$$\{\downarrow man * Name(\Delta 1, Sergey) : x1,$$

$$\downarrow man * Name(\Delta 1, Andrey) : x2\},$$

being a substring of the string *Semrepr*1 in the Example 1 is to be replaced by the string

$$(\downarrow group * Elements(\Delta 2, (\downarrow man * Name(\Delta 1, Sergey) : \{x1\},$$

$$\wedge \downarrow man * Name(\Delta 1, Andrey) : \{x2\}))$$

Let's illustrate some expressive possibilities of restricted S-languages of type 4 defined in [54].

Example 4. Let T3 be the discourse "Peter said that he had studied both in the Moscow Institute of Civil Engineering (MICE) and in the Moscow Institute of Electronic Engineering (MIEE). It was new for Somov that Peter had studied in the Moscow Institute of Electronic Engineering." The discourse T4 is associated in the Ph.D. dissertation [54] with the following semantic representation *Semrepr*2:

$$(Say[time - gramm, past](Subject, \downarrow person(Name, Peter) : \{x1\})$$

$$(Proposition1, Study1[time - gramm, past](Subject, x1)$$

$$(Learn - institution, (\downarrow techn - univer(Title, MICE) : \{x2\} \wedge$$

$$(\downarrow techn - univer(Title, MIEE) : \{x3\})) : P1)$$

$$\wedge(New(\downarrow person(Surname, Somov) : \{x4\}, P1)$$

$$\equiv Study1[time - gramm, past](Subject, x1)(Learn - institution,$$

$$(\downarrow techn - univer(Title, MICE) : \{x2\}))).$$

The analysis of this formal expression enables us to notice that the distinguished features of the proposed approach to modeling communication in NL are the possibilities listed below:

- to build (on the semantic level) the formal analogues of the phrases with indirect speech;
- to construct the compound designations of the notions and, as a consequence, the compound designations of concrete objects;
- to associate the marks with the compound descriptions of the objects (the substrings : $\{x1\}$, : $\{x2\}$, : $\{x3\}$, : $\{x4\}$);
- to associate the marks with the semantic representations of the phrases and larger fragments of a discourse (the indicator of an association of the kind in the string *Semrepr2* is the substring : *P1*);
- to build the semantic representations of the discourses with the references to the meanings of phrases and larger fragments of the considered discourse.

The class of restricted S-languages of type 4 introduced in [54] allows also for building an improved SR of the discourse T3 which the following formula *Semrepr3* :

$$(Say[time - gramm, past](Subject, \downarrow person(Name, Peter) : \{x1\})$$

$$(Time, t1)(Proposition1, (Study1[time - gramm, past](Subject, x1)(Time, t2)$$

$$(Learning - institution, (\downarrow techn - univer(Title, MICE) : \{x2\})$$

$$\wedge Study1[time - gramm, past](Subject, x1))(Time, t3)$$

$$(Learn - institution, \downarrow techn - univer(Title,$$

$$MIEE) : \{x3\})) : P1)$$

$$\wedge(New(\downarrow person(Surname, Somov : \{x4\}, P1)$$

$$\equiv Study1[time - gramm, past](Subject, x1)(Learn - institution,$$

$$x2))(Time, t2) \wedge Precedes1(t2, t1) \wedge Precedes1(t3, t1)$$

$$\wedge Precedes1(t1, current - moment)).$$

In comparison with SR *Semrepr2*, the representation *Semrepr3* is more exact, because it introduces the mark $t1$ for the short time interval of speaking by Peter, the marks $t2$ and $t3$ for time intervals when Peter had studied in the first and second university, respectively, and shows that $t2$ and $t3$ precede $t1$, and $t1$ precedes the current moment.

Example 5. Let T4 = "Somebody didn't turn off a knife-switch. This caused a fire." Then the string *Semrepr*4 of a restricted S-language of type 4

$$(\neg Switch - off[time - gramm, past](Subject, \downarrow person)(Qbject1,$$

$$\downarrow knife - switch) :: \{e1\} \wedge Cause(e1, \downarrow fire : \{e2\}))$$

may be interpreted as a semantic representation of T4 [54]. In this string, the substrings $e1$, $e2$ denote the events, and the symbol :: is used for associating events with semantic representations of assertions. The formula

$$Switch - off[time - gramm, past](Subject, \downarrow person)$$

$$(Object1, \downarrow knife - switch)$$

is built from the components $Switch - off[time, past]$ (called a predicator element in the SCL-theory), $Subject$, $\downarrow person$, $Qbject1$, $\downarrow knife - switch$ by means of applying to these components exactly one time one of the inference rules introduced in [54]. The items $Subject$ and $Object1$ are the designations of thematic roles (or conceptual cases, or semantic cases, or deep cases).

The papers [55, 56] contain the detailed proposals aimed at making more compact the complicated structure of the mathematical model constructed in [54]. It must be noted that these proposals modify a little the structure of formulas built in accordance with some rules of constructing semantic representations of NL-texts.

Example 6. In the paper [55], the discourse T4 = "Somebody didn't turn off a knife-switch. This caused a fire" is associated with the following semantic representation *Semrepr*5 :

$$(\neg Switch - off[time - gramm, past](Subject, \downarrow person)$$

$$Qbject1, \downarrow knife - switch) :: e1 \wedge Cause(e1, \downarrow fire : e2)).$$

We can see that, in accordance with the proposals from [55, 56], we use as the marks of subformulas describing the events not the strings $\{e1\}$ and $\{e2\}$ but the strings $e1$ and $e2$. In the semantic representations constructed in the Examples 1, 4, and 6, the symbol :: is used for associating the marks of events with the semantic images of statements describing these events.

It is easy to see that the symbol :: is used in the SCL-theory with the same purpose as the episodic operator $**$ in Episodic Logic [130–132, 183]. However, it was done in the year 1982, i.e., 7 years before the publication of the paper [183], where the episodic operator $**$ was introduced.

Example 7. Let T5 be the discourse "Victor said that he had lived in Kiev and Moscow. It was new for Rita that Victor had lived in Kiev." Then we can associate with T5 a semantic representation *Semrepr*6 [55] being the formula

$$(Say[time - gramm, past](Subject, \downarrow person(Name, Victor) : x1)$$

$$(Proposition, Live[time - gramm, past](Subject, x1)(Location,$$

$$(Kiev \wedge Moscow)) \; : \; P1) \wedge (New(\downarrow person(Name, Rita) \; : \; x2, P1)$$

$$\equiv Live[time - gramm, past](Subject, \downarrow person(Name, Victor) \; : \; x1)$$

$$(Location, Kiev))).$$

The string *Semrepr*6 contains the substrings : $x1$, : $x2$ but not the substrings : $\{x1\}$ and : $\{x2\}$ which could be expected by us as a consequence of our acquaintance with the semantic representation *Semrepr*2 in Example 4.

Example 8. Let T6 = "How many students are there in the Lomonosov Moscow State University?". Then the formula

$$??(Number1(all\ person * Study1[time - gramm, present](Subject, \Delta 1)$$

$$(Learning - institution, Lomonosov - Moscow - State - Univ) = ?x1)$$

may be considered as a semantic representation of the question T6 [55].

It must be added that the proposals formulated in [55, 56] concern not only 14 partial operations of building semantic representations of NL-texts stated in [54] but also two more operations schematically outlined in [53].

2.4 A Model of a Correspondence Between NL-Texts and Their Semantic Representations

The next component of Integral Formal Semantics of NL is a mathematical model elaborated in [54] and describing a correspondence between the NL-texts (sentences and discourses) and their semantic representations being the strings of restricted S-languages of type 4. This model proposes a unified description of at least two different sublanguages of NL. The first one is a collection of the texts in Russian, English, German, and some other languages expressing the commands to fulfill certain actions.

For instance, the first sublanguage contains the command C1 = "Turn to the left. The radius – 3 m." The expression C1 can be interpreted as a command to a radio-controlled model of ship. The second sublanguage contains, for example, the command C2 = "Draw two circles. The centers are the points (9, 14) and (12, 23). The diameters – 8 and 12 cm."

The general feature of these sublanguages is that they contain the commands to create one or several entities of a particular kind. The model contains the parameter *entity − sort*, its value is a semantic unit (called a sort) qualifying the class of entities to be created in accordance with the input sequence of commands.

If *entity − sort = event*, the model describes the commands to fulfill certain actions. If *entity − sort = geom − object*, the model describes the commands to draw certain geometrical figures on the plane.

The correspondence between NL-texts and their semantic representations determined by the model is based on the following central idea: the command "Turn to

the left" is replaced by the statement "It is necessary to turn to the left," the command "Draw two circles" is replaced by the statement "It is necessary to draw two circles," and so on. This approach provides the possibility to construct a semantic representation of an input discourse as a conjunction of the semantic representations of the statements.

Example 1. Let C3 = "Turn to the left and give the light signal. The radius of the turn is 3 m." Then, following [54], we can associate C3 with the following semantic representation:

$$(necessary[time-gramm, present](Goal, (turn(Orientation, left) :: \{e1\})$$

$$\wedge produce1(Result1, \downarrow signal (Kind-signal, light) : \{e2\})))$$

$$\wedge(Radius(e1) \equiv 3/m)).$$

Example 2 [54]. Consider C4 = "Draw two circles. Diameters – 8 cm and 12 cm." Then the following string is a possible semantic representation of C4:

$$(necessary[time-gramm, present] (Goal, draw(Object-geom,$$

$$\downarrow circle : \{x1, x2\})) \wedge (Diameter((x1 \wedge x2)) \equiv (8/cm \wedge 12/cm))).$$

The constructed mathematical model was later used as the theoretical basis for designing the software of a prototype of an NL-interface to a computer training complex destined for acquiring (by ship captains and their deputies) the skills necessary for preventing the collisions of ships [58, 59, 61].

2.5 The Theory of T-Calculuses and T-Languages

The theory of T-calculuses and T-languages (the TCL-theory) is an expansion of the theory of S-calculuses and S-languages (the SCL-theory). The outlines of the TCL-theory can be found in [55, 56]. This part of IFS studies in a formal way the semantic structure of the discourses defining a new notion or introducing a new designation of an object and, as a consequence, playing the role of an order to an intelligent system to include a new designation of a notion or of an object into the inner conceptual system. T-languages allow for describing semantic structure of sentences and discourses.

Example 1 [55]. Let T1 = "A tanker is a vessel for carrying liquid freights." Then there is a T-language containing the string *Semrepr*1 of the form

$$(tanker \Longleftarrow\uparrow transp) \wedge (tanker \equiv vessel*$$

$$Destination((\Delta2, Carry1(Objects, diverse\, freight * Kind(\Delta1, liquid))))$$

being a possible semantic representation of the definition T1.

Let's presume that an applied intelligent system can semantically analyze the strings of T-languages. Then the substrings with the symbols $\Longleftarrow$ or $\longleftarrow$ are to be interpreted as the commands to up date the considered knowledge base (KB). In particular, the substring $(tanker \Longleftarrow \uparrow transp)$ is to signify that an intelligent system should include in its knowledge base the notion's designation *tanker*, qualifying a transport means.

Example 2. Consider the text T2 = "Let M be the intersection of lines AH and BE, P be the intersection of lines CD and FK. Then it is necessary to prove that the line MP is a tangent to the circle with the center N and the radius 12 mm."

One can define such a T-language L_t that L_t will include the string *Semrepr*2 of the form

$$(M \longleftarrow geom - ob) \wedge (P \longleftarrow geom - ob) \wedge Intersection(AH, BE, M) \wedge$$

$$Intersection(CD, FK, P) \wedge Necessary(Prove(Proposition,$$

$$Tangent(MP, \downarrow circle(Center, N)(Radius, 12/mm)))).$$

In the string *Semrepr*2, the substrings $(M \longleftarrow geom - ob)$ and $(P \longleftarrow geom - ob)$ indicate that an intelligent system will include in its knowledge base the constants *M* and *P*, denoting some geometrical objects.

The rules allowing us to construct the formulas *Semrepr*1 and *Semrepr*2 are explained in [55]. Thus, the theory of S-calculuses and S-languages and the theory of T-calculuses and T-languages provided already in 1983 a broadly applicable variant of discourses' dynamic semantics.

The examples considered above show that the expressive power of S-languages and T-languages is very high and essentially exceeds, in particular, the expressive power of Discourse Representation Theory.

2.6 The Initial Version of the Theory of K-Calculuses and K-Languages

The SCL-theory and the TCL-theory became the starting point for developing the theory of K-calculuses, algebraic systems of conceptual syntax, and K-languages (the KCL-theory) that are nowadays the central component of IFS. The first variant of this theory elaborated in 1985 is used in [58] and is discussed in [60, 61].

The second variant is set forth in the textbook [62] and in [63, 65, 67, 68] (see also the bibliography in [65]). We'll discuss below the second variant of the KCL-theory. The basic model of the KCL-theory describes a discovered collection consisting of 14 partial operations on the conceptual structures associated with NL-texts and destined for building semantic representations of sentences and discourses.

The KCL-theory provides much more powerful formal means for describing the sets and *n*-tuples, where $(n > 1)$, than the SCL-theory. It should be noted that the KCL-theory allows for regarding the sets containing the sets and the *n*-tuples with

components being sets. This enables us to consider the relationships between the sets, untraditional functions with arguments and/or values being sets, etc.

The KCL-theory gives the definition of a class of formulas providing the possibility to (a) describe structured meanings of complicated sentences and discourses and (b) build the representations of diverse cognitive structures.

Example. Let D1 be the discourse "The chemical action of a current consists in the following: for some solutions of acids (salts, alkalis), by passing an electrical current across such a solution one can observe isolation of the substances contained in the solution and laying aside these substances on electrodes plunged into this solution. For example, by passing a current across a solution of blue vitriol ($CUSO4$) pure copper will be isolated on the negatively charged electrode. One uses this to obtain pure metals" [65].

Then D1 may have a semantic representation *Semreprdisc*1 of the form

$$(Description(action * (Kind, chemical), current1,$$

$$\exists x1\,(solution1 * (Subst, (acid \lor salt \lor alkali)))$$

$$If-then(Pass(\langle Agent1, .:\ current1\ :\ y1\rangle,$$

$$\langle Envir, x1\rangle), Observe((.:\ isolation2 * (Agent2,$$

$$diverse\,substance * Contain(x1,\#)\ :\ z1) \land$$

$$.:\ laying-aside * (Agent, z1)(Loc1,$$

$$certain\,electrode * (Plunge, x1))))\ :\ P1 \land$$

$$Example(P1, If-then(Pass(\langle Agent1, .:\ current1\ :\ y2\rangle,$$

$$\langle Envir, .:\ solution1 * (Subst,$$

$$blue-vitriol * (Formula, CuSO4))\rangle),$$

$$Isolate2(\langle Agent2, .:\ matter1 * (Is, copper * (Kind, pure))\rangle,$$

$$\langle Loc1, .:\ electrode * (Charge, neg)\rangle)))) \land$$

$$Use(.:\ phenomenon * (Charact, P1),$$

$$Obtain(*, diverse\,metal * (Kind, pure)))).$$

Here the referential structure of D1 is reflected with the help of variables $x1$, $y1$, $y2$, $z1$, $P1$; the symbol . : is interpreted as the referential quantifier, i.e., as the informational unit corresponding to the word certain.

The text D1 is taken from the textbook on physics destined for the pupils of the eighth class in Russia (the initial class – 6-year-old children – has the number 1, the last class – the number 11). This textbook was written by A. Pyoryshkin and N. Rodina and published in Moscow in 1989. This information is reflected by the K-string *Semreprdisc*2 of the form

$$.:\ text * (Content, Semreprdisc1)(Source, .:\ text-book *$$

$$(Educ-inst, any\,school * (Country, Russia)(Grade\,1, 8))$$

$$(Area, physics)(City, Moscow)(Year, 1989))$$

$$(Authors, (A.Pyoryshkin \wedge N.Rodina)) : inf\,218,$$

where the string $inf\,218$ is a mark of a concrete informational object. So we see that K-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.

Numerous examples of K-strings are adduced in [62, 65, 67, 68]. Hence the expressive power both of standard K-languages and of S-languages of type 5 considerably exceeds the expressive possibilities of other approaches to the formalization of NL-semantics discussed above.

In [65], some opportunities of recording NL-communication by means of standard K-languages are explained. That is, it is shown how it is possible to represent in a formal manner the actions carried out by intelligent systems in the course of communication.

The paper [65] also shows how to use standard K-languages for describing semantic-syntactic information associated with words and fixed word combinations.

2.7 The Theory of K-Representations as the Kernel of the Current Version of Integral Formal Semantics

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 in both Russian and English, in particular, in [65–100].

The first basic constituent of the theory of K-representations is the theory of SK-languages (standard knowledge languages), stated, in particular, in [70–94]. The kernel of the theory of SK-languages is a mathematical model describing a system of such 10 partial operations on structured meanings (SMs) of natural language texts (NL-texts) that, using primitive conceptual items as "blocks," we are able to build SMs of arbitrary NL-texts (including articles, textbooks) and arbitrary pieces of knowledge about the world. The outlines of this model can be found in two papers published by Springer in the series "Lecture Notes in Computer Science" [83, 86].

A preliminary version of the theory of SK-languages – the theory of restricted K-calculuses and K-languages (the RKCL-theory) – was set forth in [70].

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

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

The second basic constituent of the theory of K-representations is a broadly applicable mathematical model of a linguistic database (LDB). 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, participle, 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 K-representations is formed by two complex, strongly structured algorithms carrying out semantic-syntactic analysis of texts from some practically interesting sublanguages of NL. The first algorithm is described in Chapters. 8 and 9 of the book [85]. The second algorithm being a modification of the first one is set forth in Chapters 9 and 10 of this monograph. Both algorithms are based on the elaborated formal model of an LDB.

The other components of the theory of K-representations are briefly characterized in Sect. 2.2.

Problems

1. What are the components of Integral Formal Semantics of Natural Language?
2. What is a formal system or a calculus?
3. What are the new features of the theory of S-calculuses and S-languages (the SCL-theory) in comparison with the first-order predicate logic?
4. Discover the new ways of using the logical connectives $\wedge$ and $\vee$ in the SCL-theory in comparison with the first-order predicate logic.
5. What are the main ideas of building compound representations of notions (concepts) in the SCL-theory?
6. What is the purpose of using the symbols ":" and "::" in the formulas of the SCL-theory?
7. What is common for the SCL-theory and Episodic Logic?
8. What is the purpose of using the symbols $\leftarrow$ and $\Leftarrow$ in the formulas of the theory of T-calculuses and T-languages?