Chapter 5
A Study of the Expressive Possibilities of SK-Languages

Abstract In this chapter we will continue the analysis of the expressive possibilities of SK-languages. The collection of examples considered above doesn't demonstrate the real power of the constructed mathematical model. That is why let's consider a number of additional examples in order to illustrate some important possibilities of SK-languages concerning the construction of semantic representations of sentences and discourses and describing the pieces of knowledge about the world. The advantages of the theory of SK-languages in comparison, in particular, with Discourse Representation Theory, Episodic Logic, Theory of Conceptual Graphs, and Database Semantics of Natural Language are set forth. If the string *Expr* of a certain SK-language is a semantic representation of a natural language expression *T*, the string *Expr* will be called *a possible K-representation (KR) of the expression T*.

5.1 A Convenient Method of Describing Events

The key role in the formation of sentences is played by the verbs and lexical units which are the derivatives of verbs – the participles, gerunds, and verbal nouns, because they express the various relations between the objects in the considered application domain.

In computer linguistics, a conceptual relation between a meaning of a verbal form and a meaning of a word combination (or a separate word) depending on this verbal form is called *a thematic role (or conceptual case, deep case, semantic case, semantic role).*

In such different languages as Russian, English, German, and French, it is possible to observe the following regularity: in the sentences with the same verb mentioning an event, the different quantity of the thematic roles connected with a meaning of this verb is explicitly realized.

For example, let T1 = "Professor Novikov arrived yesterday" and T2 = "Professor Novikov arrived yesterday from Prague." Then in the sentence T1 two thematic roles are explicitly realized, these roles can be called *Agent1 (Agent of action)* and *Time*.

V.A. Fomichov, *Semantics-Oriented Natural Language Processing*, IFSR International Series on Systems Science and Engineering 27, DOI 10.1007/978-0-387-72926-8_5,

Meanwhile, in the sentence T2 three thematic roles *Agent1, Time, and Place 1* (the latter designates the relation connecting an event of moving in space and an initial spatial object) are explicitly realized.

Let's consider a flexible way of constructing semantic representations of the event descriptions taking into account this phenomenon of natural language. For this purpose it is required to formulate a certain assumption about the properties of the considered conceptual basis B.

Assumption 1.

The set of sorts $St(B)$ includes the distinguished sort *sit* ("situation"); the set of variables $V(B)$ includes such countable subset $Vsit = \{el, e2, e3, \ldots\}$ that for every $v \in Vsit$, $tp(v) = sit$; the primary informational universe $X(B)$ includes such binary relational symbol *Situation* that its type $tp(Situation)$ is the string of the form $\{(sit, \uparrow sit)\}$.

The meaning of the expression $\uparrow sit$ in the right part of the relationship

$$tp(Situation) = \{(sit, \uparrow sit)\}$$

is as follows: we will be able to build the expressions of the form

$$Situation(e_k, concept\_descr),$$

where e_k is a variable denoting an event (a sale, a purchase, a flight, etc.) and *concept_descr* is a simple or compound denotation of a notion being a semantic characteristic of the event.

Let's agree that a connection between a mark of a situation and a semantic description of this situation will be given by means of the formulas of the kind

$$Situation(var, cpt * (rel_1, d_1) \ldots (rel_n, d_n)),$$

where *var* is a variable of type *sit*, the element *cpt* belongs to $X(B)$ and is interpreted as a notion qualifying a situation, $n \geq 1$, for $k = 1, \ldots, n$, rel_k is a characteristic of the considered situation, d_k is the value of the characteristic rel_k.

Example. Let the expressions $Expr1$ and $Expr2$ be defined as follows:

$$Expr1 = \exists e1\,(sit)\,(Situation(e1, arrival * (Time, x1)$$

$$(Agent1, person * (Qualif, professor)(Surname,$$

$$'Novikov') : x2)) \wedge Before(x1, \#now\#)),$$

$$Expr2 = \exists el\,(sit)\,(Situation(e1, arrival * (Time, x1)$$

$$(Agent1, person * (Qualif, professor)(Surname, 'Novikov') : x2)$$

$$(Place1, city * (Name1, 'Prague') : x3)) \wedge Before(x1, \#now\#)).$$

Then it is easy to see that it is possible to construct such conceptual basis B that Assumption 1 will be true, and the following relationships will take place:

$$B(0, 1, 2, 3, 4, 5, 7, 8, 9) \Rightarrow Exprl, Expr2 \in Ls(B),$$

$$Exprl \,\&\, prop \in Ts(B),\ Expr2 \,\&\, prop \in Ts(B),$$

where $prop = P(B)$ is the distinguished sort "a meaning of proposition" of the conceptual basis B.

This method of describing the events will be used many times below. Most often, for the sake of compactness, the existential quantifier followed by a variable marking-up an event will be omitted. For example, instead of the formula $Expr1$, we will consider the following formula $Expr3$ of the form

$$(Situation(e1, arrival * (Time, x1)(Agent1, person*$$

$$(Qualif, professor)(Surname, 'Novikov') : x2)) \wedge Before(xl, \#now\#)).$$

5.2 Formalization of Assumptions About the Structure of Semantic Representations of Sets

The statements, questions, commands can include the designations of sets. In order to have a unified approach to constructing semantic representations of sets' descriptions, it is reasonable to formulate a number of additional assumptions about the considered conceptual bases.

With respect to the fact that the designations of sets in texts often include the designations of natural numbers ("five three-tonne containers", etc.), we will suppose that the following requirements are satisfied for the considered conceptual basis B :

Assumption 2.

The set of sorts $St(B)$ includes the distinguished sort *nat* ("natural number"), the primary informational universe $X(B)$ includes a subset of strings Nt consisting of all strings of the form $d_1 \ldots d_k$, where $k \geq 1$, for $m = 1, \ldots, k$, d_m is a symbol from the set

$$\{'0', 1, '2', '3', '4', '5', '6', 7, '8', '9'\},$$

and it follows from $d_1 = '0'$ that $k = 1$. Besides, for each z from the set Nt, $tp(z) = nat$.

Let's also demand that the primary information universe $X(B)$ includes the distinguished elements

$$set, Numb, Qual - compos, Object - compos$$

interpreted as follows: *set* is the designation of the concept "a finite set," *Numb* is the name of an one-argument function "The quantity of elements of a set," $Qual - compos$ is the name of the binary relation "Qualitative composition of a set," $Object - compos$ is the name of the binary relation "Object composition of a set."

Assumption 3.
The primary informational universe $X(B)$ includes such elements *set*, *Numb*, $Qual - compos$, $Object - compos$ that

$$tp(set) = \uparrow \{[entity]\},$$

$$t(Numb) = \{(\{[entity]\}, nat)\},$$

$$tp(Qual - compos) = \{(\{[entity]\}, [concept])\},$$

$$tp(Object - compos) = \{(\{[entity]\}, [entity])\},$$

where $[entity]$, $[concept]$, $[object]$ are the basic types "entity," "concept," "object" (see Chap. 3).

Let's consider the interpretation of the distinguished elements of the primary informational universe $X(B)$ mentioned in the Assumption 3.

Using the elements *set*, *Numb*, and any string *number* from Nt, we can construct a semantic representation of the expression "a certain set containing *number* elements" in the form

$$certn\, set * (Numb, number),$$

where $certn = ref(B)$ is the referential quantifier of the considered conceptual basis B.

The purpose of considering the binary relational symbol $Qual - compos$ is as follows. Let v be a variable designating a certain set, and *conc* be a simple or compound designation of a notion. Then the expression

$$Qual - compos(v, conc)$$

designates the meaning of the statement "Each element of the set v is qualified by the notion *conc*," and this statement can be true or false. The examples of such expressions are the formulas

$$Qual - compos(S1, container1),$$

$$Qual - compos(S2, paper1),$$

$$Qual - compos(S3, container1 * (Material, aluminum)),$$

$$Qual - compos(S4, paper1 * (Area1, biology)).$$

On the other hand, the symbol $Qual - compos$ will be also employed in constructing the compound designations of the sets in the form

$$ref\, set * (Qual - compos, conc) : v,$$

where ref is the the referential quantifier of the considered conceptual basis B; *conc* is a simple or compound designation of a notion (a concept), v is a variable.

In particular, the conceptual basis B can be chosen so that the language Ls (B) includes the expressions

$$certn\,set * (Qual - compos, container1) : S1,$$

$$certn\,set * (Qual - compos, paper1) : S2,$$

$$certn\,set * (Qual - compos, container1 * (Material, aluminum)) : S3,$$

$$certn\,set * (Qual - compos, paper1 * (Area1, biology)) : S4.$$

A fragment of a text designating a set can be an explicit enumeration of the elements of this set. For instance, the text "Two customers, the joint-stock company 'Rainbow' and the Open Company 'Zenith,' have not paid September deliveries" contains a fragment of the kind.

The binary relational symbol $Object - compos$ is intended, in particular, for constructing the expressions of the form

$$Object - compos(v, (x_1 \wedge x_2 \ldots \wedge x_n)),$$

where $v, x_1, x_2, \ldots, x_n$ are the variables, v designates a set, and $x_1, x_2, \ldots, x_n$ are the designations of all elements belonging to the set with the designation v.

For example, we'll consider the expression

$$(Object - compos(y1, (x1 \wedge x2)) \wedge$$

$$Is1(x1, joint - stock - comp) \wedge Is1(x2, open - comp) \wedge$$

$$Name(x1, \text{``}Rainbow\text{''}) \wedge Name(x2, \text{``}Zenith\text{''}))$$

as a semantic representation of the statement "The set $y1$ consists of the joint-stock company 'Rainbow' and the Open Company 'Zenith'," where the first company is designated by $x1$, and the second company has the designation $x2$.

At the same time we should have the opportunity (if necessary) to build the formulas with the structure

$$ref\,set * (Object - compos, (x_1 \wedge x_2 \wedge \ldots, x_n)) : y1,$$

where ref is the referential quantifier, and $x_1 \ldots, x_n, y_1$ are the variables of type $[entity]$ (the basic type "entity").

For instance, we should have the opportunity of constructing the expression

$$certn\,set * (Object - compos, (x1 \wedge x2)) : y1.$$

Example. Let $Setdescr1$ and $Setdescr2$ be the expressions "3 containers with ceramics from India" and "a party consisting of the boxes with the numbers 3217, 3218, 3219" respectively.

Then it is possible to construct such conceptual basis B that the Assumption 2 will be true, and $Ls(B)$ will include the formula

$$certn\,set * (Numb, 3)(Qual - compos, container1*$$
$$(Content1, certn\,set * (Qual - compos, manufact\_product*$$
$$(Kind, ceramics)(Country, India))))$$

and the formula

$$certn\,set * (Numb, 3)(Object - compos, (certn\,box1*$$
$$(Number1, 3217) : x1 \wedge certn\,box1 * (Number1, 3218) : x2$$
$$\wedge\ certn\,box1 * (Number1, 3219) : x3)) : S1.$$

The constructed formulas will be interpreted as possible K-representations of the expressions *Setdescr*1 and *Setdescr*2; here *xl*, *x*2, *x*3 are the labels of boxes, *S*1 is the label of a set.

5.3 Semantic Representations of Questions with the Role Interrogative Words

Let *Interrog_expr* be the designation of the union of the set of all interrogative adverbs and the set of all word groups composed by interrogative pronouns and by interrogative pronouns with associated prepositions. Then it is possible to distinguish in the set *Interrog_expr* a subset including, in particular, the words and word groups "who," "to whom," "from whom," "when," "where." In order to formulate the property of each element of this subset, we shall introduce the designation *nil* for an empty preposition. If an interrogative pronoun *qswd* is used in any question without preposition, let's agree to say that this pronoun *qswd* is associated in this question with the empty preposition *nil*.

Therefore, for each pronoun *qswd* from the considered subset, there is such preposition *prep* that the pair $(prep, qswd)$ corresponds to a certain thematic role. Rather often, there are several prepositions *prep* satisfying this condition.

For example, the pairs

$$(to, whom), (for, whom), (from, whom)$$

correspond to the thematic roles *Addressee*, *Addressee*, *Source*1, respectively. We see that the different pairs (to, who), (for, who) correspond to one thematic role *Addressee*, and it is confirmed by the analysis of the phrases "Who is the book sent to?" and "Who is the book sent for?" The thematic role "Source1" is realized, in particular, in the sentence "This book was sent by Yves."

The pronouns and adverbs belonging to the specified subset will be called below *the role interrogative words.*

Assumption 4.
The primary informational universe $X(B)$ of the conceptual basis B includes the symbol *Question*, and

$$tp(Question) = \{([entity], P)\},$$

where $tp = tp(B)$ is a mapping associating a type with an informational unit, $[entity]$ is the basic type "entity," $P = P(B)$ is the distinguished sort "a meaning of proposition."

Let Assumption 4 be true for the conceptual basis B. Then the semantic representation of a question with n role interrogative words can be presented in the form

$$Question(v_1, A)$$

in case $n = 1$ and in the form

$$Question((v_1 \wedge \ldots \wedge v_n), A)$$

in case $n > 1$, where $v_1 \ldots v_n$ are the variables, and A is an l-formula depending on the variables $v_1 \ldots v_n$ and displaying the content of a statement (i.e. being a semantic representation of a statement).

Example 1. Let Qs1 be the question "Where has the three-ton aluminum container arrived from?" and *Exprl* be the string

$$Question(x1, Situation(el, receipt1 * (Time, certn\,moment*$$

$$(Before, \#now\#) : t1)(Location1, x1)(Object1,$$

$$certn\,container1 * (Weight, 3/ton)(Material, aluminum) : x2)))\,.$$

Then it is easy to construct such conceptual basis B that Assumption 1 and Assumption 4 are true for B, $P(B) = prop$, and

$$B(0, 1, 2, 3, 4, 5, 8) \Rightarrow Expr1 \in Ls(B),$$

$$Expr1\,\&\,prop \in Ts(B).$$

The expression $Expr1$ is a possible K-representation of the question Qs1. In this expression, the symbols xl, $x2$, $e1$, $t1$ are the variables, $receipt1$ is the informational unit (in other words, semantic unit) corresponding to the noun "receipt" and expressing the meaning "transporting a physical object to a spatial object" (unlike the meaning "receipt of a Ph.D. degree").

Example 2. Let Qs2 = "When and where has the three-ton aluminum container arrived from?" Then a K-representation of the question Qs2 can be the expression

$$Question((t1 \wedge x1), Situation(el, receipt1 * (Time, certn\,moment*$$

$$(Before, \#now\#) : t1)(Place1, x1)(Object1,$$

$$certn\,container1 * (Weight, 3/ton)(Material, aluminum) : x2)))\,.$$

5.4 Semantic Representations of Questions About the Quantity of Objects and Events

Assumption 5.

Let $X(B)$ be a conceptual basis of the form (S, Ct, Ql), where S is a sort system, Ct is a concept-object system, and Ql is a system of quantifiers and logical connectives of the form

$$(ref, int1, int2, eq, neg, binlog, ext),$$

and let $X(B)$ include such elements $arbitrary$, all, $Elem$ that

$$tp(arbitrary) = int1, \, tp(all) = int2,$$

$$tp(Elem) = \{([entity], \{[entity]\})\}.$$

The elements $arbitrary$, all, $Elem$ are interpreted as informational units "any" ("arbitrary"), "all," and "An element of a set" (the name of the relation "To be an element of a set").

It is necessary to notice that $int1$ and $int2$ are the distinguished elements of the set of sorts $St(B)$. By definition (see Sect. 3.8), the elements $int1$ and $int2$ are the types of intensional quantifiers from the first and second considered classes respectively.

Example 1. Let Qs1 = "How many copies of the books by P.N. Somov are available in the library?" Then it is possible to define such conceptual basis B that Assumptions 4 and 5 are true for B, and the expression

$$Question(x1, (x1 \equiv Numb(all\, copy1 * (Inform-object,$$

$$arbitrary\, book * (Authors, certn\, person * (Initials, 'P.N.')$$

$$(Surname, 'Somov') : x2) : x3)(Storage-place, certn\, library : x4))))$$

belongs to the SK-language $Ls(B)$. Therefore, this expression is a possible K-representation of the question Qs1.

Example 2. Let Qs2 = "How many people participated in the creation of the textbook on statistics?", then a possible K-representation of the question Qs2 can have the form

$$Question(x1, ((x1 \equiv Numb(all\, person * (Elem, S1))) \wedge$$

$$Description(arbitrary\, person * (Elem, S1) : y1, (Situation(e1,$$

$$participation1 * (Agent1, y1)(Time, x2)(Kind-of-activity,$$

$$creation1 * (Product1, certn\, textbook * (Area1, statistics)))) \wedge$$

$$Before(x2, \#Now\#))))).$$

Example 3. Let Qs3 = "How many books did arrive in January of this year to the library No. 18?" Then the formula

$$Question(x1, ((x1 \equiv Numb(all\,book * (Elem, S1))) \wedge$$
$$Description(arbitrary\,book * (Elem, S1) : y1,$$
$$Situation(e1, receipt2 * (Object1, y1)$$
$$(Time, \langle 01, \#current\_year\# \rangle)(Place2,$$
$$certn\,library * (Number1, 18) : x2)))))$$

is a possible K-representation of the question Qs3.

Example 4. The question Qs4 = "How many times did Mr. Stepan Semyonov fly to Mexico?" can have the following possible K-representation:

$$Question(x1, (x1 \equiv Numb(all\,flight * (Agent1,$$
$$certn\,person * (Name, 'Stepan')(Surname, 'Semyonov') : x2)$$
$$(Place2, certn\,country * (Name1, 'Mexico') : x3)$$
$$(Time, arbitrary\,moment * (Before, \#now\#))))).$$

5.5 Semantic Representations of Questions with an Interrogative Pronoun Attached to a Noun

The method proposed above for constructing K-representations of the questions with the role interrogative words can also be used for building K-representations of the questions with the pronoun "what" attached to a noun in singular or plural.

Example 1. Let Qs1 = "What publishing house has released the novel 'The Winds of Africa'?" Then the string

$$Question(x1, (Situation(e1, releasing1 * (Time, x2)$$
$$(Agent2, certn\,publish\_house : x1)(Object, certn\,novel1*$$
$$(Name1, 'The\,Winds\,of\,Africa') : x3)) \wedge Before(x2, \#now\#)))$$

can be interpreted as a K-representation of the question Qs1.

Example 2. Let Qs2 = "What foreign publishing houses the writer Igor Nosov is collaborating with?" Then the formula

$$Question(S1, (Qual-compos(S1, publish\_house*$$
$$(Kind-geogr, foreign)) \wedge Description(arbitrary\,publish\_house*$$
$$(Elem, S1) : y1, Situation(e1, cooperation1 * (Agent,$$
$$certn\,person * (Profession, writer)(First\_name, 'Igor')$$

$$(Surname, 'Nosov') : x1)(Organization1, y1)(Time, \#now\#)))))$$

is a possible K-representation of the question Qs2.

5.6 Semantic Representations of General Questions

The questions with the answer "Yes/No" are called in linguistics *general questions*. The form of representing the meaning of questions with the interrogative words proposed above can also be used for constructing semantic representations of general questions. With this aim, each question of the kind will be interpreted as a request to specify the truth value of a certain statement. For example, it is possible to interpret the question Qs1 = "Is Gent a city of Belgium?" as a request to find the truth value of the statement "Gent is one of the cities of Belgium." For realizing this idea, we introduce

Assumption 6.

The set of sorts $St(B)$ includes the distinguished sort *boolean* called "logical value"; the primary informational universe $X(B)$ includes the different elements $truth$, $false$, and $tp(truth) = tp(false) = boolean$; the set of functional symbols $F(B)$ includes such unary functional symbol $Truth-value$ that

$$tp(Truth-value) = \{(P, boolean)\},$$

where $P = P(B)$ is the distinguished sort "a meaning of proposition."

Example 1. Suppose that for the considered conceptual basis B the Assumption 6 is true. Then a K-representation of the question Qs1 = "Is Gent a city of Belgium?" can have the form

$$Question(x1, (x1 \equiv Truth-value(Elem(certn\,city*$$

$$(Name, \text{“}Gent\text{”}) : x2, all\,city * (Part1,$$

$$certn\,country * (Name, \text{“}Belgium\text{”}) : x3))))).$$

In this formula, the symbol *certn* is the referential quantifier $ref(B)$ of the conceptual basis B; $xl, x2, x3$ are the variables from $V(B)$.

Example 2. Let Qs2 = "Did the international scientific conference 'COLING' take place in Asia ?" Then the formula

$$Question(xl, (x1 \equiv Truth-value((Situation(el, holding1*$$

$$(Event1, certn\,conf * (Type1, intern)(Type2, scient)$$

$$(Name, 'COLING') : x2)(Place, certn\,continent*$$

$$(Name, \text{“}Asia\text{”}) : x3)(Time, x4)) \wedge Before(x4, \#now\#)))))$$

can be interpreted a K-representation of the question Qs2.

5.7 Describing Semantic Structure of Commands

Let's proceed from two basic ideas. First, when we speak about a command or about an order, we always mean that there is one intelligent system forming the command (it is designated by the expression $\#Operator\#$) and another intelligent system (or a finite set of intelligent systems) which should execute the command or an order (it is designated by expression $\#Executor\#$). Secondly, a verb in an imperative mood or the infinitive form of a verb will be replaced by the corresponding verbal noun.

Assumption 7.

The set of sorts $St(B)$ includes the distinguished elements $ints$ (the sort "intelligent system"), mom (the sort "a moment of time"); the primary informational universe $X(B)$ includes such elements $Command$, $\#Operator\#$, $\#Executor\#$, $\#now\#$ that

$$tp = (Command) = \{(ints, ints, mom, \uparrow sit)\},$$

$$tp(\#Operator\#) = tp(\#Executor\#) = ints,$$

$$tp(\#now\#) = mom.$$

Example. Let Comm1 = "Deliver a box with the details to the warehouse No. 3," where Comm1 is the command transmitted by the operator of a flexible industrial system to an intelligent transport robot. Let $Semrepr1$ be the K-string

$$Command(\#Operator\#, \#Executor\#, \#now\#,$$

$$delivery1 * (Object1, certn\, box * (Content1,$$

$$certn\, set * (Qual - compos, detail1)) : x1)$$

$$(Place2, certn\, warehouse * (Number1, 3) : x2))\ .$$

Then the basis B can be defined in such a way that Assumptions 1 and 7 will be true, and the following relationship will take place:

$$B(0, 1, 2, 3, 4, 5, 8) \Rightarrow Semrepr1 \in Ls(B).$$

The K-string $Semrepr1$ can be interpreted as a K-representation of the command Comm1.

5.8 Representation of Set-Theoretical Relationships and Operations on Sets

Example 1. Let T1a be the sentence "Namur is one of the cities of Belgium." Then we shall proceed from the semantically equivalent text T1b = "Namur belongs to set of all cities of Belgium" in order to construct a K-representation of the sentence T1a as the K-string $Expr1$ of the form

$$Elem(certn\, city * (Name1, \text{“}Namur\text{”}) : x1,$$

$$all\, city * (Part1, certn\, country * (Name1, \text{“}Belgium\text{”}) : x2)) .$$

Then there is such conceptual basis B that

$$B(0,1,2,3,8,1,5,0,1,2,3,8,1,5,2,4) \Rightarrow Expr1 \in Ls(B).$$

While constructing the K-string $Expr1$, the following assumptions were used:

$$Elem, country, city, certn, Part1 \in X(B),$$

$$tp(Elem) = \{([entity], \{[entity]\})\},$$

$$tp(country) = tp(city) = \uparrow space.object;$$

$$tp(Part1) = \{(space.object, space.object)\},$$

and $certn = ref(B)$ is the referential quantifier of the conceptual basis B.

Example 2. Let T2a be the command = “Include the container No. 4318 into the party to be sent to Burgas.” We shall transform the command T2a into the statement T2b = “An operator has ordered to include the container No. 4318 into a certain party sent to the city Burgas.”

Then it is possible to construct a K-representation of the texts T2a and T2b in the form

$$Command(\#Operator\#, \#Executor\#, \#Now\#, inclusion1*$$

$$(Object1, certn\, container1 * (Number1, 4318) : x1)(Target-set,$$

$$certn\, party2 * (Place-destin, certn\, city*$$

$$(Name, \text{“}Burgas''): x2) : S1)),$$

where $S1$ is the label of a party of production. It is possible to present in a similar way the orders about the division of set of objects into certain parts and about the assembly of several sets into one, as in case of overloading the details from several boxes into one box.

5.9 Semantic Representations of Phrases with Subordinate Clauses of Purpose and Indirect Speech

Example 1. Let 1 = “Alexander entered the State University – Higher School of Economics (HSE) in order to acquire the qualification ‘Business Informatics,’ ” and let $Semrepr1$ be the K-string

$$(Situation(e1, entering2 * (Agent1, certn\, person*$$

$$(Name, \text{“}Alexander\text{”}) : x1)(Lern\_institution, certn\, university*$$

$$(Name1, \text{“}HSE\text{”}) : x2)(Time, t1)(Purpose,$$
$$acquisition1 * (New\_property, certn\, qualification1 * (Name1,$$
$$business\_informatics) : x3))) \wedge Before(t1, \#now\#)).$$

Then there is such conceptual basis B that

$$B(0, 1, 2, 3, 4, 5, 7, 8) \Rightarrow Semrepr1 \in Ls(B)\ ,$$
$$Semrepr1\,\&\, prop \in Ts(B)\ .$$

Example 2. Let 2 = “The director said that the reorganization of the firm is planned on February,” and let $Semrepr2$ be the K-string

$$(Situation(e1, oral\_communication * (Agent1,$$
$$Director(certn\, org : xl))(Time, t1)(Content2, Planned(t1,$$
$$certn\, reorganization * (Object\_org, certn\, firm1 : x1),$$
$$Nearest(February, t1)))) \wedge Before(t1, \#now\#)).$$

Then it is easy to construct such conceptual basis B that the following relationships take place:

$$B(0, 2) \Rightarrow Nearest(February, t1) \in Ls(B)\ ,$$
$$Nearest(February, t1)\,\&\, time\_interval \in Ts(B)\,;$$
$$B(0, 1, 2, 4, 5, 7, 8) \Rightarrow Semrepr2 \in Ls(B)\ ,$$
$$Semrepr2\,\&\, prop \in Ts(B)\ .$$

5.10 Explicit Representation of Causal and Time Relations in Discourses

As it was mentioned above, a discourse, or a coherent text, is a sequence of the phrases (complete or incomplete, elliptical) with interconnected meanings. The *referential structure of an NL-text* is a correspondence between the groups of words from this text and the things, events, processes, meanings designated by these groups of words. The SK-languages provide broad possibilities of describing semantic structure of discourses, in particular, their referential structure.

Example. Let T1 = “The first-year student Peter Gromov didn’t notice that the schedule had changed. As a result, he missed the first lecture on linear algebra.” In this text, we can observe the following feature of discourses: the personal pronoun “he” is used instead of the longer combination “The first-year student Peter

Gromov." One says that the last expression and the pronoun "he" have the same referent being a certain person, a student of a college or a university.

Obviously, in order to specify the referent structure of a text, it is necessary to connect the labels with the entities designated by the groups of words from this text or implicitly mentioned in the text. We shall make this in the following way:

- the implicitly mentioned educational institution – the label $x1$;
- "The first-year student Peter Somov," "he" – the label $x2$;
- "the schedule" – the label $x3$;
- "the first lecture on linear algebra " – the label $x4$;
- "didn't notice" – the label $e1$ (the event 1);
- "has changed" – the label $e2$;
- $e3$ – the label of the situation described by the first sentence of T1;
- "missed" – the label $e4$ (event).

We shall suppose that a semantic representation of the text T1 should include the fragment $Cause\ (e3, e4)$.

Let $Semrepr1$ be the K-string

$$((Situation(e1, \neg noticing1 * (Agent1, certn\, person*$$

$$(Name, \text{"}Peter''\text{)}(Surname, \text{"}Gromov\text{"})(Qualif, student*$$

$$(Year1, 1)(Learn\_institution, x1)) : x2)(Time, t1)$$

$$(Object\_of\_attention, e2)) \wedge Before(e1, \#now\#) \wedge$$

$$Situation(e2, change1 * (Object1, certn\, schedule : x3)(Time, t2))$$

$$\wedge Before(t2, t1)) : P1 \wedge Characterizes(P1, e3)) .$$

Then $Semrepr1$ is a possible K-representation of the first sentence $S1$ of the discourse T1.

Let $Semrepr2$ be the K-string

$$(Situation(e4, missing1 * (Agent1, x2)(Event1,$$

$$certn\, lecture1 * (Discipline, linear - algebra)$$

$$(Learn\_institution, x1) : x4)(Time, t3))$$

$$\wedge Before(t3, \#now\#)).$$

Then $Semrepr2$ can be interpreted as a possible K-representation of the second sentence $S2$ from the discourse T1. Let

$$Semdisc1 = (Semrepr1 \wedge Semrepr2 \wedge Cause(e3, e4)) .$$

Then the formal expression $Semdisc1$ is a possible semantic representation of the discourse T1 being its K-representation.

5.11 Semantic Representations of Discourses with the References to the Meanings of Phrases and Larger Parts of the Text

Example. Let T1 = "The join-stock company 'Rainbow' will sign the contract till December 15th. The deputy director Igor Panov has told this." Here the pronoun "this" designates the reference to the meaning of the first sentence of the discourse T1. Let $Semrepr1$ be the expression

$$(Situation(e1, signing1 * (Agent2, certn\, organization*$$

$$(Type, joint\_stock\_comp)(Name1, \text{“}Rainbow\text{”}) : x1)(Time, t1)$$

$$(Inf\_object, certn\, contract1 : x2))\wedge$$

$$Before(t1, Date(12, 15, \#current\_year\#))),$$

and $Semrepr2$ be the expression of the form

$$(Semrepr1 : P1 \wedge Situation(e2, oral\_message * (Agent1,$$

$$certn\, person * (Name, \text{“}Igor\text{”})(Surname, \text{“}Panov\text{”}) : x3)(Time, t2)$$

$$(Content2, P1)) \wedge Before(t2, \#now\#)\wedge$$

$$Deputy\_Director(x3, certn\, organization : x4)).$$

Then there is such conceptual basis B that

$$B(0, 1, 2, 3, 4, 5, 7, 8) \Rightarrow Semrepr2 \in Ls(B)\,,$$

$$Semrepr2\,\&\, prop \in Ts(B)\,,$$

where $prop = P(B)$ is the distinguished sort "a meaning of proposition."

The rule $P[5]$ allows for attaching a variable v to a semantic representation $Semrepr$ of arbitrary narrative text and for obtaining the formula

$$Semrepr : v\,,$$

where v is a variable of the sort $P(B)$ ("a meaning of proposition").

Therefore, the variables of the sort $P(B)$ will be the images of the expressions "about this," "this method," "this question," etc., in the complete semantic representation of the considered discourse (in the same way as in the last example).

5.12 Representing the Pieces of Knowledge About the World

Example 1. Let T1 = "The notion 'a molecule' is used in physics, chemistry and biology." It is possible to define such conceptual basis B that the set of sorts $St(B)$

includes element $activity\_field$, the primary informational universe $X(B)$ includes the elements

$$activity\_field,\ string,\ notion,\ \text{“}molecule\text{”},\ Use,\ Notion-name,$$

$$physics,\ chemistry,\ biology,$$

and the types of these elements are set by the relationships

$$tp(notion) = [\uparrow concept], tp(\text{“}molecule\text{”}) = string,$$

$$tp(physics) = tp(chemistry) = tp(biology) = activity\_field,$$

$$tp(Use) = \{([concept],\ activity\_field)\},$$

$$tp(Notion-name) = \{([concept],\ string)\}.$$

Let $certn$ be the referential quantifier, Use and $Notion-name$ be the binary relational symbols not being the names of the functions, and

$$s1 = Notion-name(certn\, notion,\ \text{“}molecule\text{”}),$$

$$s2 = notion * (Notion-name,\ \text{“}molecule\text{”}),$$

$$s3 = Use(certn\, notion * (Notion-name,\ \text{“}molecule\text{”}),$$

$$(physics \wedge chemistry \wedge biology)).$$

Then $B(0, 1, 4) \Rightarrow s1 \in Ls(B)$; $B(0, 1, 4, 8) \Rightarrow s2 \in Ls(B)$;

$$B(0, 1, 4, 8, 1, 0, 7, 4) \Rightarrow s3 \in Ls(B).$$

The built formula $s3$ is a possible semantic representation of the sentence T1.

Example 2. Let T2 be the definition “Teenager is a person at the age from 12 to 19 years”; $semdef$ be the K-string

$$((teenager \equiv person * (Age,\ x1)) \wedge$$

$$\neg Less1(x1,\ 12/year) \wedge \neg Greater1(x1,\ 19/year)).$$

Then $semdef$ is a possible K-representation of T2.

5.13 Object-Oriented Representations of Knowledge Pieces

We can build complex designations of objects and sets of objects, using SK-languages.

Example 1. We can build the following K-representation of a description of the international scientific journal “Informatica”:

$$certn\,int.sc.journal * (Title, \text{"}Informatica\text{"})$$

$$(Country, Slovenia)(City, Ljubljana)(Fields,$$

$$(artif.intel \wedge cogn.science \wedge databases)) : k225\,,$$

where $k225$ is the mark of the knowledge module with the data about "Informatica."

The definition of the class of SK-languages allows for building the formal conceptual representations of texts as informational objects reflecting not only the meaning but also the external characteristics (or metadata) of the text: the authors, the date, the application fields of described methods and models, etc.

Example 2. In a way similar to the way used in the previous example, we can construct a knowledge module stating the famous Pythagorean Theorem and also indicating its author and field of science. For instance, such a module may be the following expression of a certain SK-language:

$$certn\,textual\_object * (Kind1, theorem)(Fields1,$$

$$geometry)(Authors, Pythagoras)(Content\_inf\_ob,$$

$$\exists x_1\,(geom)\,\exists x_2\,(geom)\,\exists x_3\,(geom)\,\exists x_4\,(geom)$$

$$If-then((Is1(x_1, right-triangle) \wedge$$

$$Hypotenuse(x_2, x_1) \wedge Leg1((x_3 \wedge x_4), x_1)),$$

$$(Square(Length(x_2)) \equiv Sum(Square(Length(x_3)),$$

$$Square(Length(x_4)))))) : k81.$$

5.14 The Marked-Up Conceptual Bases

The analysis shows that it is possible and reasonable to select a compact collection of primary informational units to be used for constructing semantic representations of NL-texts independently on application domains, in particular, for building SRs of questions, commands, and descriptions of sets.

The Assumptions 1–7 formulated above indicate such primary informational units. The purpose of introducing the definitions below is to determine the notion of a marked-up conceptual basis, i.e., a conceptual basis satisfying the Assumptions 1–7.

Definition 5.1. Let B be an arbitrary conceptual basis, $St(B)$ be the set of sorts of the basis B, $P(B)$ be the sort "a meaning of proposition," $X(B)$ be the primary informational universe of the basis B. Then a system Qmk of the form

$$(\,sit, Vsit, Situation, Question, boolean, true, false, Truth-value\,) \tag{5.1}$$

is called *a marking-up of questions of the conceptual basis B* $\Leftrightarrow$ when

- the elements sit, $boolean$ belong to the set-theoretical difference of the set of sorts $St(B)$ and the set $Concr(P)$, consisting of all sorts being the concretizations of the sort $P(B)$ with respect to the generality relation $Gen(B)$,
- $X(B)$ includes the different elements $Situation$, $Question$, $boolean$, $true$, $false$, $Truth-value$,
- Assumptions 1, 4, and 6 are true for the components of this system.

Definition 5.2. Let B be an arbitrary conceptual basis. Then a system $Setmk$ of the form

$$(nat, Nt, set, Numb, Qual-compos, Object-compos, arbitrary, all, Elem) \tag{5.2}$$

is called *a set-theoretical marking-up of the conceptual basis B* $\Leftrightarrow$ when

- the element nat belongs to the set-theoretical difference of the set of sorts $St(B)$ and the set $Concr(P)$, consisting of all sorts being the concretizations of the sort $P(B)$ with respect to the generality relation $Gen(B)$,
- Nt is a subset of the primary informational universe $X(B)$,
- the elements set, $Numb$, $Qual-compos$, $Object-compos$, $arbitrary$, all, $Elem$ are different elements of the set $X(B)$,
- Assumptions 2, 3, and 5 are true for the components of this system.

Definition 5.3. Let B be an arbitrary conceptual basis, Qmk be a marking-up of questions of the form (5.1) of the basis B. Then a system Cmk of the form

$$(\,ints, mom, \#now\#, \#Operator\#, \#Executor\#, Command\,) \tag{5.3}$$

is called *a marking-up of commands of the basis B coordinated with the marking-up of questions Qmk* $\Leftrightarrow$ when

- $ints$, mom, $\#now\#$, $\#Operator\#$, $\#Executor\#$, $Command$ are different elements of the set $X(B)$,
- $ints, mom \in St(B) \setminus (Concr(P) \cup Concr(sit) \cup \{boolean\})$, where $Concr(P)$ and $Concr(sit)$ are the sets of all sorts being respectively the concretizations of the sort $P(B)$ and of the sort sit with respect to the generality relation $Gen(B)$,
- Assumption 7 is true for the components of the system Cmk.

The formal notions introduced above enable us to make the final step and to join these notions in the definition of the class of marked-up conceptual bases.

Definition 5.4. *A marked-up conceptual basis (m.c.b.)* is an arbitrary four-tuple Cb of the form

$$(\,B, Qmk, Setmk, Cmk\,)\,, \tag{5.4}$$

where B is an arbitrary conceptual basis, Qmk is a marking-up of questions of the form (5.1) for the basis B, Qmk is a set-theoretical marking-up of the conceptual basis B, Cmk is a marking-up of commands of the basis B coordinated with the marking-up of questions Qmk, and the following conditions are satisfied:

- all components of the systems Qmk, $Setmk$, Cmk, except for the component Nt of the system $Setmk$, are the different elements of the primary informational universe $X(B)$;
- if $Stadd = \{sit, boolean, ints, mom, nat\}$ and $Concr(P)$ is the set of all sorts being the concretizations of the sort $P(B)$ ("a meaning of proposition") with respect to the generality relation $Gen(B)$, then $Stadd$ is a subset of the set $St(B) \setminus Concr(P)$, and every two different elements of the set $Stadd$ are incomparable both for the generality relation $Gen(B)$ and for the tolerance relation Tol;
- if s is an arbitrary element of the set $Stadd$, and u is an arbitrary element of the set $Concr(P)$, then the sorts s and u are incomparable for the tolerance relation Tol.

Definition 5.5. Let's agree to say that a marked-up conceptual basis Cb is a marked-up basis of the standard form $\Leftrightarrow$ Cb is a system of the form (5.4), Qmk is a system of the form (5.1), $Setmk$ is a system of the form (5.2), and Cmk is a system of the form (5.3).

We'll consider below the marked-up conceptual bases of only standard form.

The class of formal languages

$$\{Ls(B) \mid B \text{ is the first component of arbitrary m.c.b. } Cb\}$$

will be used as the class of semantic languages while describing the correspondences between NL-texts and their semantic representations.

5.15 Related Approaches to Representing Semantic Structure of NL-Texts

Proceeding from the ideas stated in this chapter, one is able to easily simulate the expressive mechanisms provided by Discourse Representation Theory (DRT), Episodic Logic, and Theory of Conceptual Graphs (TCG).

For instance, the example in Sect. 5.10 shows how it would be possible to describe causal and time relationships between the events mentioned in a discourse without the episodic operator.

The manner of building short compound (though rather simple) designations of the notions and sets of objects proposed by TCG can be replaced by much more general methods of the theory of K-representations.

Example 1. Let T = "Sue sent the gift to Bob," and a possible TCG-representation of T1 is

$$[Person : Sue] \leftarrow (Agent) \leftarrow [Send] \rightarrow$$
$$(Theme) \rightarrow [gift] : \#]$$
$$(Recipient) \rightarrow [Person : Bob].$$

Then a possible K-representation of T1 is

$$Situation(e1, sending1 * (Time, certn\, mom*$$

$$(Before, \#now\#) : t1)Agent, certn\, person*$$

$$(Name, \text{"}Sue\text{"}) : x1)(Theme, certn\, gift : x2)$$

$$(Recipient, certn\, person * (Name, \text{"}Bob\text{"}) : x3))\,.$$

Example 2. A certain group of vehicles can be denoted in TCG by the expression

$$[vehicle : \{*\}]$$

and by the K-string

$$certn\, set * (Qual - compos, vehicle)\,.$$

Example 3. Let's agree that the string *at* designates a special symbol used in e-mail addresses. Then a certain set consisting of 70 books can be denoted in TCG by the expression

$$[book : \{*\}\, at\, 70]$$

and by the K-string

$$certn\, set * (Numb, 70)(Qual - compos, book)\,.$$

Example 4. A set consisting of three concrete persons with the names Bill, Mary, and Sue can have a TCG-representation

$$[person : \{Bill, Mary, Sue\}]$$

and a K-representation

$$certn\, set * (Object - compos, (certn\, person*$$

$$(Name, \text{"}Bill\text{"}) \wedge certn\, person * (Name, \text{"}Mary\text{"})$$

$$\wedge\, certn\, person * (Name, \text{"}Sue\text{"})))\,.$$

Example 5. The set of all farmers located in the state of Maine can have a TCG-representation

$$[\,[Farmer : \lambda] \to (Loc) \to [State : Maine] : \{*\}\forall]$$

and a K-representation

$$all\, farmer * (Loc, certn\, state2*$$

$$(Name1, \text{"}Maine\text{"}))\,.$$

The material of this chapter helps also to understand how it would be possible to simulate other expressive mechanisms of TCG, in particular, the manners to represent the finite sequences and the semantic structure of the sentences with direct or indirect speech.

The principal advantage of the theory of SK-languages is that the convenient ways of simulating numerous expressive mechanisms of NL are the specific combinations of discovered general operations on conceptual structures.

Due to this feature, there is no necessity, as in TCG, to invent special combinations of symbols (for instance, $\{*\}$ or $\{*\}$ *at* 70 or $\{*\}\forall$) as the indicators of special constructions of semantic level.

The concrete advantages of the theory of SK-languages in comparison with first-order logic, Discourse Representation Theory (DRT), and Episodic Logic (EL) are, in particular, the possibilities to

1. distinguish in a formal way objects (physical things, events, etc.) and notions qualifying these objects;
2. build compound representations of notions;
3. distinguish in a formal manner the objects and the sets of objects, the notions and the sets of notions;
4. build complicated representations of sets, sets of sets, etc.;
5. describe set-theoretical relationships;
6. effectively describe structured meanings (SMs) of discourses with the references to the meanings of phrases and larger parts of discourses;
7. describe SMs of sentences with the words "concept," "notion";
8. describe SMs of sentences where the logical connective "and" or "or" joins not the expressions – assertions but the designations of things, sets, or notions;
9. build complex designations of objects and sets;
10. consider non traditional functions with arguments and/or values being the sets of objects, of notions, of texts' semantic representations, etc.;
11. construct formal analogues of the meanings of infinitives with dependent words and, as a consequence, to represent proposals, goals, obligations, and commitments.

It should be added that the model has at least three global distinctive features as regards its structure and destination in comparison with EL.

The first feature is as follows: In fact, the purpose of this monograph is to represent in a mathematical form a hypothesis about the general mental mechanisms (or operations) underlying the formation of complicated conceptual structures (or semantic structures, or knowledge structures) from basic conceptual items.

EL doesn't undertake an attempt of the kind, and 21 Backus-Naur forms used in [129] for defining the basic logical syntax rather disguise such mechanisms (operations) in comparison with the more general 10 rules described in Chap. 4.

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

The third global distinction is that the form of describing in EL the basic collection of informational units is not a strictly mathematical one.

For example, the collection of Backus-Naur forms used with this purpose in [129] contains the expressions

$$\langle l-place-pred-const\rangle ::= happy \mid person \mid certain \mid probable \mid \ldots,$$

$$\langle 1-fold-pred-modifier-const > ::=$$
$$plur \mid very \mid \mid former \mid almost \mid in-manner \mid \ldots.$$

The only way to escape the use of three dots in productions is to define an analogue of the notion of a conceptual basis introduced in this monograph.

The items (4)–(8), (10), (11) in the list above indicate the principal advantages of the theory of SK-languages in comparison with the Theory of Conceptual Graphs(TCG). Besides, the expressive possibilities of the new theory are much higher than the possibilities of TCG as regards the items (1), (2), (9).

There are numerous technical distinctions between the theory of SK-languages and Database Semantics of Natural language [119, 120]. The principal global distinction is that the theory of SK-languages puts forward (in a mathematical form) a hypothesis about the organization of structured meanings associated not only with separate sentences in NL but also with arbitrary complex discourses.

On the contrary, the Database Semantics of NL doesn't propose the formal tools being convenient for studying the problem of formalizing semantic structure of complex long discourses.

It is the same principal distinction and principal advantage as in the case of the comparison with the approach to representing semantic structure of NL-sentences and short discourses with the help of the language UNL (Universal Networking Language). A number of concrete advantages of SK-languages in comparison with UNL, in particular, concerning the representation of complex concepts (called scopes) is analyzed in [84, 85, 90, 93, 94].

Problems

1. Describe the components of a marking-up of questions of a conceptual basis.
2. Explain the assumptions about the types of distinguished informational units

$$set, Numb, Qual-compos, Object-compos, arbitrary, all, Elem$$

   being the components of a set-theoretical marking-up of a conceptual basis.
3. What is the difference between a conceptual basis and a marked-up conceptual basis?
4. How is it proposed to build semantic representations of events?
5. What are the main ideas of constructing compound semantic descriptions of the sets?
6. How is it possible to build the K-representations of commands?
7. Construct the K-representations of the following NL-texts:

(a) How many parcels from Reading have been received?
(b) Is it possible for a pensioner to get a credit?
(c) What journal published for the first time an article about BMW 330?
(d) Which scholars from Belgium did attend the conference?
(e) BMW 750 was put for sale in the year 1985.
(f) It is planned to inaugurate next year the offices of this bank in Omsk and Tomsk.
(g) Dr. William Jones, the president of the university, visited in March the University of Heidelberg (Germany). This was written in the newsletter of the University of Heidelberg.