Chapter 4
A Mathematical Model for Describing Structured Meanings of Natural Language Sentences and Discourses

Abstract The purpose of this chapter is to construct a mathematical model describing a system consisting of ten partial operations on the finite sequences with the elements being structured meanings of Natural Language (NL) expressions. Informally, the goal is to develop a mathematical tool being convenient for building semantic representations both of separate sentences in NL and of complex discourses of arbitrary big length pertaining to technology, medicine, economy, and other fields of professional activity. The starting point for developing this model is the definition of the class of conceptual bases introduced in the previous chapter. The constructed mathematical model includes the definition of a new class of formal systems, or calculuses – the class of K-calculuses (knowledge calculuses) and the definition of a new class of formal languages – the class of SK-languages (standard knowledge languages).

4.1 The Essence of a New Approach to Formalizing Semantic Structure of Natural Language Texts

The analysis shows that the task of modeling numerous expressive mechanisms interacting in NL goes far beyond the scope of first-order predicate logic and beyond the scope of popular approaches to the formalization of NL semantics. That is why it seems to be reasonable to develop an original formal approach to this problem, starting from a careful consideration of the fundamental presuppositions underlying a formalism to be elaborated.

4.1.1 Toward Expanding the Universe of Formal Study

Let's imagine that we want to investigate a problem with the help of formal means and, with this aim, to consider a set of entities (real and abstract) and, besides, some relations with the attributes being the elements of these sets and some functions

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_4,

with the arguments and values from this set. Then let's call such a set of entities *a universe of formal study*.

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

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

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

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

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

These are just the unique features possessed by the theory of SK-languages (see Fig. 4.1). In particular, this theory allows for including in the universe of formal study the following elements of new kinds:

- a compound designation of a notion

$$scholar * (Field\_of\_knowledge, biology)(Degree, Ph.D.)\,;$$

- a compound designation of a goal of a young scholar

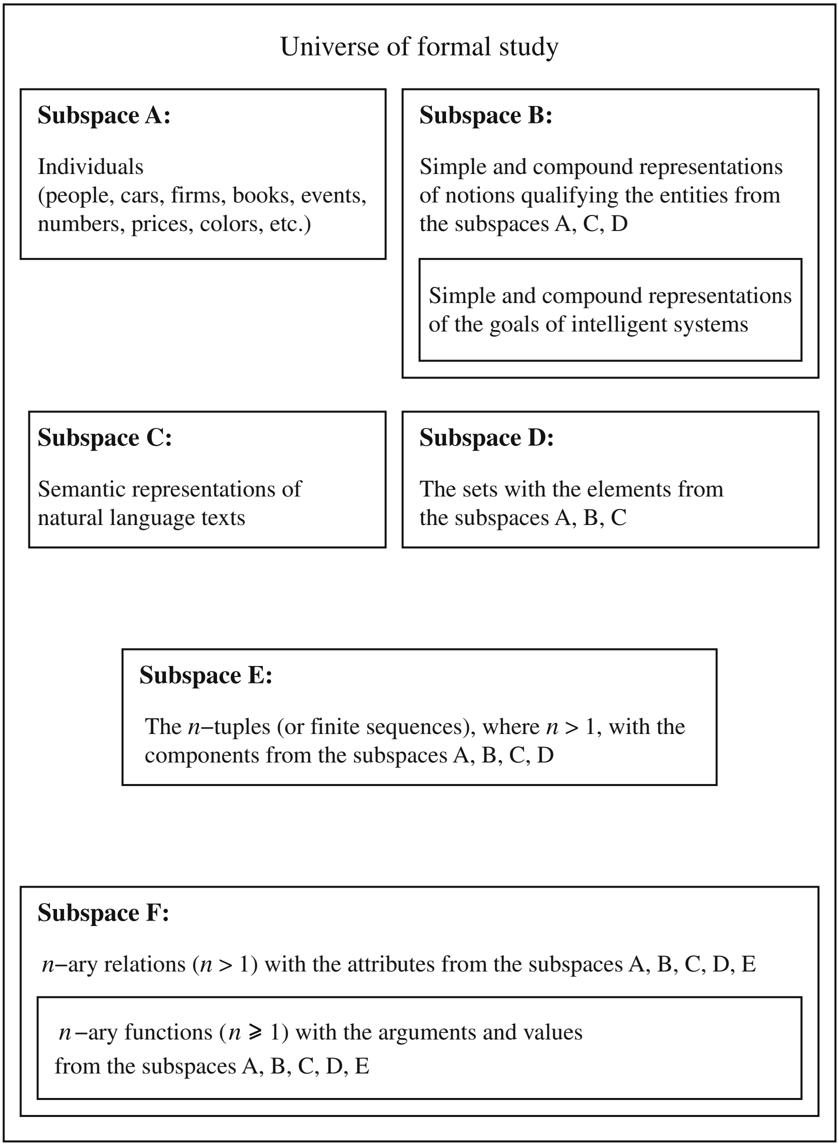

Fig. 4.1 The structure of the expanded universe of formal study

$$Defending2 * (Sci - institution, Stanford - University)$$
$$(Kind\_of\_dissertation, Ph.D.dissertation*$$
$$(Field\_of\_knowledge, computer\_science));$$

- a compound designation of a set

$$certain\ art - collection * (Quantity, 17)$$

$$(Qualitative\_composition, vase).$$

4.1.2 The Algebraic Essence of the Model Describing Conceptual Operations

During last decade, the most popular approaches to building formal representations of the meanings of NL-texts have been Discourse Representation Theory (DRT) [143, 144], Theory of Conceptual Graphs (TCG), represented, in particular, in [195, 196], and Episodic Logic (EL) [130–132, 182, 184]. In fact, DRT and TCG are oriented at describing the semantic structure of only sentences and short simple discourses. EL studies the structure of only a part of discourses, more exactly, of discourses where the time and causal relationships between the situations (called episodes) are realized.

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

The definition of the class of restricted standard knowledge languages (RSK-languages) [70, 76] became the first mathematically complete answer in English to the following question: how would it be possible to describe in a mathematical way a system of operations on conceptual structures allowing for building (after a finite number of steps) semantic representations (SRs) of arbitrarily complicated sentences and discourses from arbitrary application domains, starting from primary informational items.

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

Expanding this approach to studying semantics of NL, let's consider the main ideas of determining a new class of formal languages called SK-languages. Our starting point will be the definition of the class of conceptual bases introduced in the preceding chapter.

Each conceptual basis B determines three classes of formulas, the first class $Ls(B)$ being considered as the principal one and being called *the SK-language (standard knowledge language) in the stationary conceptual basis* B. Its strings (they are called K-strings) are convenient for building SRs of NL-texts. We'll consider below only the formulas from the first class $Ls(B)$.

In order to determine for arbitrary conceptual basis B three classes of formulas, a collection of some rules $P[0], P[1], \ldots, P[10]$ for building well-formed expressions is defined. The rule $P[0]$ provides an initial stock of formulas from the first class. For example, there is such a conceptual basis B that, according to $P[0]$, $Ls(B)$ includes the elements

$$container1,\ blue,\ country,\ France,\ set,\ 12,\ all,\ arbitrary,$$

$$Height,\ Distance,\ Quantity,\ Authors,\ Friends,\ Suppliers,\ x1,\ x2,\ e3,\ P7.$$

For arbitrary conceptual basis B, let $Degr(B)$ be the union of all Cartesian m-degrees of $Ls(B)$, where $m \geq 1$. Then the meaning of the rules of constructing well-formed formulas $P[0], P[1], \ldots, P[10]$ can be explained as follows: for each k from 1 to 10, the rule $P[k]$ determines a partial unary operation $Op[k]$ on the set $Degr(B)$ with the value being an element of $Ls(B)$.

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

$$\langle \wedge, Belgium, The-Netherlands, Luxemburg \rangle$$

is the K-string

$$(Belgium \wedge The-Netherlands \wedge Luxemburg).$$

Thus, the essence of the basic model of the theory of SK-languages is as follows: this model determines a partial algebra of the form

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

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

4.1.3 Shortly About the Rules for Building Semantic Representations of Natural Language Texts

It was mentioned above that the goal of introducing the notion of conceptual basis is to get a starting point for constructing a mathematical model describing (a) the regularities of structured meanings both of separate sentences and of complex discourses in NL; (b) a collection of the rules allowing for building semantic representations both of sentences and complex discourses in NL, starting from the primary units of conceptual level and using a small number of special symbols.

Let's consider now the basic ideas underlying the definitions of the rules intended for building semantic representations of sentences and complex discourses in NL. These ideas are stated informally, with the help of examples. The exact mathematical definitions can be found in the next chapter.

Let's regard (ignoring many details) the structure of strings which can be obtained by applying any of the rules $P[1], \ldots, P[10]$ at the last step of inferring the formulas. The rule $P[1]$ enables us to build K-strings of the form *Quant Conc*, where *Quant* is a semantic item corresponding to the meanings of such words and expressions as "certain," "any," "arbitrary," "each," "all," "several," etc. (such semantic items will be called *intensional quantifiers*), and *Conc* is a designation (simple or compound) of a concept. The examples of K-strings for $P[1]$ as the last applied rule are as follows:

$$certn\,container1,\ all\,container1,$$

$$certn\,consignment,$$

$$certn\,container1 * (Content1,\ ceramics),$$

where the last expression is built with the help of both the rules $P[0]$, $P[1]$ and the rule with the number 4, the symbol *certn* is to be interpreted as the informational item corresponding to the expression "a certain."

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

$$Distance(Moscow,\ Paris),$$

$$Weight(certn\,container1 * (Color,\ blue)(Content1,\ ceramics)).$$

Using the rule $P[3]$, we can build the strings of the form $(a1 \equiv a2)$, where $a1$ and $a2$ are K-strings formed with the help of any rules from $P[0], \ldots, P[10]$, and $a1$ and $a2$ represent the entities being homogeneous in a certain sense. The following expressions are the examples of K-strings constructed as a result of employing the rule $P[3]$ at the last step of inference:

$$(Distance(Moscow,\ Paris) \equiv x1),$$

$$(y1 \equiv y3),\ (Height(certn\,container1) \equiv 2/m).$$

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

$$Belong(Bonn,\ Cities(Germany)),$$

$$Subset(certn\,series1 * (Name - origin,\ tetracyclin),\ all\,antibiotic).$$

The rule $P[5]$ enables us to construct the K-strings of the form $Expr : v$, where *Expr* is a K-string not including v, v is a variable, and some other conditions are satisfied. Using $P[5]$, one can mark by variables in the semantic representation of any NL-text: (a) the descriptions of diverse entities mentioned in the text (physical objects, events, concepts, etc.), (b) the semantic representations (SRs) of sentences and of larger texts' fragments to which a reference is given in any part of a text. The examples of K-strings for $P[5]$ are as follows:

$$certn\,container1 : x3,$$

$$Higher(certn\,container1 : x3,\ certn\,container1 : x5) : P1.$$

The rule $P[5]$ provides the possibility to form SRs of texts in such a manner that these SRs reflect the referential structure of NL-texts. This means that an SR of an NL-text includes the variables being the unique marks of the various entities mentioned in the text; the set of such entities can include the structured meanings of some sentences and larger fragments of a discourse being referred to in this discourse.

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

$$\neg antibiotic\,,$$

$$\neg Belong(penicillin,\ certn\,series1 * (Name - origin,\ tetracyclin)).$$

Using the rule $P[7]$, one can build the K-strings of the form

$$(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$$

or of the form

$$(a_1 \vee a_2 \vee \ldots \vee a_n),$$

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

$$(streptococcus \vee staphylococcus),$$

$$(Belong((Bonn \wedge Hamburg \wedge Stuttgart),\ Cities(Germany))$$

$$\wedge \neg Belong(Bonn,\ Cities((Finland \vee Norway \vee Sweden)))).$$

The rule $P[8]$ allows us to build, in particular, the K-strings of the form

$$cpt * (rel_1,\ val_1), \ldots, (rel_n,\ val_n),$$

where cpt is an informational item from the primary informational universe $X(B)$ designating a concept (a notion), for $k = 1, \ldots, n$, rel_k is the name of a function with one argument or of a binary relation, val_k designates a possible value of rel_k for objects characterized by the concept cpt in case rel_k is the name of a function, and val_k designates the second attribute of rel_k in case rel_k is the name of a binary relation. The following expressions are the examples of K-strings obtained as a result of applying the rule $P[8]$ on the final step of inference:

$$container1 * (Content1,\ ceramics),$$

$$consignment * (Quantity, 12)(Compos1,$$

$$container1 * (Content1, ceramics)).$$

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

$$\forall n1\,(integer)\exists n2\,(integer)Less(n1, n2),$$

$$\exists y\,(country * (Location, Europe))Greater(Quantity(Cities(y)), 15).$$

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

4.1.4 The Scheme of Determining Three Classes of Formulas Generated by a Conceptual Basis

Let's consider in more detail the suggested original scheme of an approach to determining three classes of well-formed expressions called formulas.

The notions introduced above enable us to determine for every conceptual basis B a set of formulas $Forms(B)$ being convenient for describing structured meanings (SMs) of NL-texts and operations on SMs.

Definition 4.1. Let B be an arbitrary conceptual basis, *Specsymbols* be the set consisting of the symbols ′,′ (comma), ′(′, ′)′, ′ : ′, ′ ∗ ′, ′⟨′, ′⟩′. Then

$$D(B) = X(B) \cup V(B) \cup Specsymbols,$$

$$Ds(B) = D(B) \cup \{'\&'\},$$

$D^+(B)$ and $Ds^+(B)$ are the sets of all non empty finite sequences of the elements from $D(B)$ and $Ds(B)$ respectively.

The essence of the approach to determining conceptual formulas proposed in this book is as follows. As stated above, some assertions $P[0], P[1], \ldots, P[10]$ will be defined; they are interpreted as the rules of building semantic representations of NL-texts from the elements of the primary universe $X(B)$, variables from $V(B)$, and several symbols being the elements of the set *Specsymbols*. The rule $P[0]$ provides an initial stock of formulas.

If $1 \leq i \leq 10$, then for an arbitrary conceptual basis B and for $k = 1, \ldots, i$, the assertions $P[0], P[1], \ldots, P[i]$ determine by conjoint induction some sets of formulas

$$Lnr_i(B) \subset D^+(B),$$

$$T^0(B), Tnr_i^1(B), \ldots, Tnr_i^i(B) \subset Ds^+(B),$$

$$Ynr_i^1(B), \ldots, Ynr_i^i(B) \subset Ds^+(B).$$

The set $Lnr_i(B)$ is considered as the main subclass of formulas generated by $P[0], \ldots, P[i]$. The formulas from this set are intended for describing structured meanings (or semantic content) of NL-texts.

If $1 \leq k \leq i$, the set $Tnr_i^k(B)$ consists of strings of the form $b \,\&\, t$, where $b \in Lnr_i(B)$, $t \in Tp(S(B))$, and b is obtained by means of applying the rule $P[k]$ to some simpler formulas at the final step of an inference. It should be added that for constructing b from the elements of $X(B)$ and $V(B)$, one may use any of the rules $P[0], \ldots, P[k], \ldots, P[i]$; these rules may be applied arbitrarily many times.

If a conceptual basis B is chosen to describe a certain application domain, then b can be interpreted as a semantic representation of a text or as a fragment of an SR of a text pertaining to the considered domain. In this case, t may be considered as the designation of the kind of the entity qualified by such SR or by a fragment of SR. Besides, t may qualify b as a semantic representation of a narrative text.

The number i is interpreted in these denotations as the maximal ordered number of such rules from the list $P[0], P[1], \ldots, P[10]$ that these rules may be employed for building semantic representations of NL-expressions or for constructing knowledge modules.

For instance, it will be shown below that the sets $Lnr_4(B_1), \ldots, Lnr_{10}(B_1)$ include the formulas

$$Elem(P.Somov, Friends(J.Price)),$$

$$Elem(Firm\_Ocean, Suppliers(Firm\_Rainbow)),$$

and the sets $Tnr_4^4(B_1), \ldots, Tnr_{10}^4(B_1)$ include the formulas

$$Elem(P.Somov, Friends(J.Price)) \,\&\, prop,$$

$$Elem(Firm\_Ocean, Suppliers(Firm\_Rainbow)) \,\&\, prop,$$

where $prop$ is the distinguished sort $P(B_1)$ ("a meaning of proposition").

Each string $c \in Ynr_i^k(B)$, where $1 \leq k \leq i$, can be represented in the form

$$c = a_1 \,\&\, \ldots \,\&\, a_m \,\&\, b,$$

where $a_1, \ldots, a_m$, $b \in Lnr_i(B)$. Besides, there is such type $t \in Tp(S(B))$ that the string $b \,\&\, t$ belongs to $Tnr_i^k(B)$.

The strings $a_1, \ldots, a_m$ are obtained by employing some rules from the list $P[0], \ldots, P[i]$, and b is constructed from "blocks" $a_1, \ldots, a_m$ (some of them could be a little bit changed) by applying just one time the rule $P[k]$. The possible quantity of "blocks" $a_1, \ldots, a_m$ depends on k. Thus, the set $Ynr_i^k(B)$ fixes the result of applying the rule $P[k]$ just one time.

For instance, we'll see below that the sets $Ynr_4^4(B_1), \ldots, Ynr_{10}^4(B_1)$ include the formulas

$$Elem \,\&\, P.Somov \,\&\, Friends(J.Price) \,\&\, Elem(P.Somov,\, Friends(J.Price)),$$

$$Elem \,\&\, Firm\_Ocean \,\&\, Suppliers(Firm\_Rainbow)$$
$$\&\, Elem(Firm\_Ocean,\, Suppliers(Firm\_Rainbow)).$$

Let for $i = 1, \ldots, 10$,

$$T_i(B) = T^0(B) \cup Tnr_i^1(B) \cup \ldots \cup Tnr_i^i(B);$$

$$Y_i(B) = Ynr_i^1(B) \cup \ldots \cup Ynr_i^i(B);$$
$$Form_i(B) = Lnr_i(B) \cup T_i(B) \cup Y_i(B).$$

We'll interpret $Form_i(B)$ as the set of all formulas generated by the conceptual basis B with the help of the rules $P[0], \ldots, P[i]$. This set is the union of three classes of formulas, the principal class being $Lnr_i(B)$. The formulas from these three classes will be called l-formulas, t-formulas, and y-formulas respectively (see. Fig. 4.2).

The class of l-formulas is needed for assigning a type from $Tp(S(B))$ to each $b \in Lnr_i(B)$, where $i = 1, \ldots, 10$. For $q = 0, \ldots, 9$, $Lnr_q(B) \subseteq Lnr_{q+1}(B)$. The set $Lnr_{10}(B)$ is called *the standard knowledge language* (or *SK-language, standard K-language*) in the stationary conceptual basis B and is designated as $Ls(B)$. That's why l-formulas will be often called also *K-strings*.

The set $T_{10}(B)$ is designated as $Ts(B)$. For every conceptual basis B and arbitrary formula A from the set $Ts(B)$, there exist such type $t \in Tp(S(B))$ and such formula $C \in Ls(B)$ that $A = C\&t$. We'll employ only the formulas from the subclasses $Ls(B)$ and $Ts(B)$ (i.e., only l-formulas and t-formulas) for constructing semantic representations of NL-texts. y-formulas are considered as auxiliary ones, such formulas are needed for studying the properties of the sets $Ls(B)$ and $Ts(B)$.

The pairs of the form $(B, Rules)$, where B is a conceptual basis, $Rules$ is the set consisting of the rules $P[0], \ldots, P[10]$, will be called *the K-calculuses (knowledge calculuses).*

4.2 The Use of Intensional Quantifiers in Formulas

The term "intensional quantifier" was introduced in Chap. 3 for denoting the conceptual items (in other words, semantic items, informational items) associated, in particular, with the words and word combinations "every," "a certain," "arbitrary," "any," "all," "almost all," "a few," "several," "many."

The set of intensional quantifiers consists of two subclasses Int_1 and Int_2. These subclasses are defined in the following way.

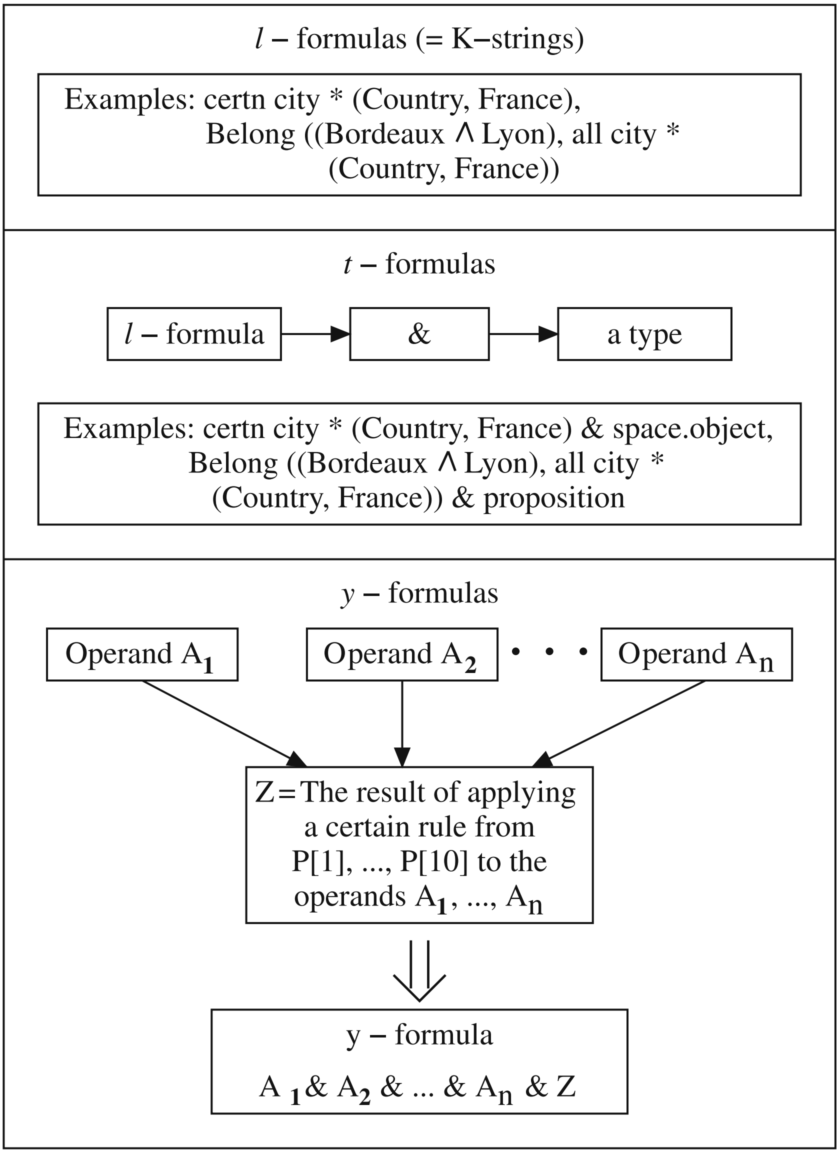

Fig. 4.2 Three classes of formulas determined by a conceptual basis

Each conceptual base B is a system of the form (S, Ct, Ql). The component Ql is a finite sequence of formal objects including, in particular, the distinguished sorts int_1 and int_2. This enables us to define the set $Int_m(B)$, where $m = 1, 2$, as the set of all such conceptual items qtr from the primary informational universe $X(B)$ that $tp(qtr) = int_m(B)$.

The elements of the set Int_1 designate, in particular, the meanings of the expressions "every," "a certain," "arbitrary," "any." The elements of the set Int_2 are interpreted as the denotations of the meanings, in particular, of the expressions "all," "almost all," "a few," "several," "many"; such words and word combinations are used for forming the designations of the sets. The minimal requirement to the set Int_1 is that this set includes the conceptual item associated with the word combination "a certain"; this conceptual item is called *the referential quantifier*. The minimal requirement to the set Int_2 is that Int_2 includes the conceptual item all.

The rule $P[1]$ allows us to join the intensional quantifiers to simple designations (a descriptor) or compound designations of the notions (concepts) . As a result of applying this rule one obtains, in particular

- the l-formulas of the form $qtr\,cpt$, where qtr is an intensional quantifier from the set $Int(B)$, cpt is a designation of a simple or compound notion;
- t-formulas of the form $qtr\,cpt\,\&\,t$, where t is a type from the set $Tp(S(B))$.

For instance, it is possible to define a conceptual basis B in such a way that, using the rule $P[0]$ during the first step of the construction process and the rule $P[1]$ during the final step, it will be possible to build:

- the l-formulas

$$certn\,city,\ certn\,city * (Name1,\ London),$$

$$every\,city,\ every\,person * (Profession,\ painter),$$

$$all\,city,\ all\,city * (Country,\ Russia),$$

- the t-formulas

$$certn\,city\,\&\,space.object,\ all\,city\,\&\,\{space.object\},$$

$$every\,person * (Profession,\ painter)\,\&\,ints * dyn.phys.ob,$$

where $ints$ is the sort "intelligent system," $dyn.phys.phys.ob$ is the sort "dynamic physical object."

Definition 4.2. If B is a conceptual basis, then for $m = 1, 2$,

$$Int_m(B) = \{qtr \in X(B) \mid tp(qtr) = int_m(B)\},$$

$$Int(B) = Int_1(B) \cup Int_2(B),$$

$$Tconc(B) = \{t \in Tp(S(B)) \mid t \text{ has the beginning } \uparrow\} \cup Spectp,$$

where

$$Spectp = \{[\uparrow entity],\ [\uparrow concept],\ [\uparrow object]\},$$

and the elements of the set $Spectp$ are interpreted as the types of informational units associated with the *words* "entity," "notion" ("concept"), "object," respectively.

Using the rules $P[0]$ and $P[l]$, we can build some strings of the form

$$quantifier\ concept\_descr\,,$$

where

$$quantifier \in Int(B),\ concept\_descr \in X(B),$$

$$tp(concept\_descr) \in Tconc(B).$$

For instance, if B_1 is the conceptual basis determined in Chap. 3, we can construct the l-formulas

$$certn\,person,\ certn\,tour.gr,\ certn\,concept,$$

$$all\,person,\ all\,tour.gr,\ all\,concept$$

generated by B_1.

It is possible also to build more complex strings of the form $qt\,cpt$, where qt is an intensional quantifier, with the help of the rule $P[1]$, using preliminary the rule $P[8]$ (see next sections) and, may be, some other rules (besides the rules $P[0]$ and $P[1]$) for constructing the string cpt denoting a concept. For instance, it will be possible to build the l-formulas

$$certain\,tour.group * (Numb,\ 12),$$

$$all\,tour.group * (Numb,\ 12)$$

in the conceptual basis B_1 constructed in Chap. 3. These formulas are interpreted as semantic representations of the expressions "a certain tourist group consisting of 12 persons" and "all tourist groups consisting of 12 persons."

The transition from an l-formula cpt designating a notion to a certain l-formula $qtr\,cpt$, where qtz is an intensional quantifier, is described with the help of a special function h.

Definition 4.3. Let B be a conceptual basis, $S = S(B)$. Then the mapping h : $\{1, 2\} \times Tp(S) \rightarrow Tp(S)$ is determined as follows: if $u \in Tp(S)$ and the string $\uparrow u$ belongs to Tp(S), then

$$h(1,\ \uparrow u) = u,\ h(2,\ \uparrow u) = \{u\};$$

$$h(1,\ [\uparrow entity]) = [entity],\ h(2,\ [\uparrow entity]) = \{[entity]\};$$

$$h(1,\ [\uparrow concept]) = [concept],\ h(2,\ [\uparrow concept]) = \{[concept]\};$$

$$h(1,\ [\uparrow object]) = [object],\ h(2, [\uparrow object]) = \{[object]\}.$$

From the standpoint of building semantic representations (SRs) of texts, the mapping h describes the transformations of the types in the course of the transition:

- from the notions "a person," "a tourist group" to the SRs of expressions "some person," "arbitrary person," "some tourist group," "arbitrary tourist group," etc. (in case the first argument of h is 1) and to the SRs of expressions "all people," "all tourist groups," etc. (if the first argument of h is 2);

- from the notions "an entity," "an object," "a concept" to the SRs of the expressions "some entity," "arbitrary entity," "some object," "arbitrary object," "some concept," "arbitrary concept," etc. (when the first argument of h is 1) and to the SRs of the expressions "all entities," "all objects," "all concepts," etc. (when the first argument of h is 2).

Definition 4.4. Denote by $P[1]$ the assertion "Let $cpt \in L(B) \setminus V(B)$, the type $u \in Tconc(B)$, $k \in \{0, 8\}$, and the string $cpt \,\&\, u$ belong to $T^k(B)$. Let $m \in \{1, 2\}$, $qtr \in Int_m$, $t = h(m, u)$, and b be the string of the form $qtr\, cpt$. Then $b \in L(B)$, the string of the form $b \,\&\, t$ belongs to $T^1(B)$, and the string of the form $qtr \,\&\, cpt \,\&\, b$ belongs to $Y^1(B)$."

Example 1. Let B be the conceptual basis B_1 constructed in Chap. 3;

$L(B)$, $T^0(B)$, $T^1(B)$, $Y^1(B)$ be the least sets of formulas jointly defined by the assertions $P[0]$ and $P[1]$. Then it is easy to verify that the following relationships take place:

$$person \in L(B) \setminus V(B),\ person \,\&\uparrow ints * dyn.phys.ob \in T^0(B),$$

$$certn \in Int_1(B),\ h(1, \uparrow ints * dyn.phys.ob) = ints * dyn.phys.ob \Rightarrow$$
$$certn\ person \in L(B),\ certn\ person \,\&\, ints * dyn.phys.ob \in T^1(B),$$
$$certn \,\&\, person \,\&\, certn\ person \in Y^1(B);$$
$$all \in Int_2(B),\ h(2, \uparrow ints * dyn.phys.ob) = \{ints * dyn.phys.ob\} \Rightarrow$$
$$all\ person \in L(B),\ all\ person \,\&\, \{ints * dyn.phys.ob\} \in T^1(B),$$
$$all \,\&\, person \,\&\, all\ person \in Y^1(B);$$
$$tour.group \in L(B) \setminus V(B),$$
$$tour.group \&\uparrow \{ints * dyn.phys.ob\} \in T^0(B),$$

$$h(1, \uparrow \{ints * dyn.phys.ob\}) = \{ints * dyn.phys.ob\},$$

$$h(2, \uparrow \{ints * dyn.phys.ob\}) = \{\{ints * dyn.phys.ob\}\} \Rightarrow$$
$$certn\, tour.group,\ all\, tour.group \in L(B),$$
$$certn\, tour.group \,\&\, \{ints * dyn.phys.ob\},$$
$$all\, tour.group \,\&\, \{\{ints * dyn.phys.ob\}\} \in T^1(B),$$
$$certn \,\&\, tour.group \,\&\, certn\, tour.gr \in Y^1(B),$$
$$all \,\&\, tour.group \,\&\, all\, tour.gr \in Y^1(B),$$

$$concept \in L(B) \setminus V(B),\ concept \,\&\, [\uparrow concept] \in T^0(B);$$

$$h(1, [\uparrow concept]) = [concept],$$

$$h(2, [\uparrow concept]) = \{[concept]\} \Rightarrow$$

$$certn\,concept,\ all\,concept \in L(B),$$

$$certn\,concept\,\&\,[concept],\ \ all\,concept\,\&\,\{[concept]\} \in T^1(B),$$

$$certn\,\&\,concept\,\&\,certn\,concept \in Y^1(B),$$

$$all\,\&\,concept\,\&\,all\,concept \in Y^1(B).$$

Comment to the rule P[1]. The fragment of the rule $P[1]$ "Let $cpt \in L(B) \setminus V(B)$, the type $u \in Tconc(B)$, $k \in \{0, 8\}$, and the string $cpt\,\&\,u$ belong to $T^k(B)$" means that u is the type of a notion, i.e. either u has the beginning $\uparrow$ or u is one of the symbols $[\uparrow entity]$, $[\uparrow concept]$, $[\uparrow object]$.

If $k = 0$, cpt designates a primitive (unstructured) concept. If $k = 1$, cpt is a compound denotation of a concept. Such compound denotations of the concepts will be constructed with the help of the rule $P[8]$; the examples of the kind are the expressions

$$person * (Activity.field,\ biology),\ city * (Country,\ Russia).$$

The rule $P[1]$ will be very often used below for constructing semantic representations of NL-texts, because it is necessary for building semantic images of the expressions formed by the nouns with dependent words. For example, let Qs1 be the question "Where has been the two-tonne green container delivered from?" Then, as a result of fulfilling the first step of constructing a SR of Qs1, it is possible to obtain the expression

$$certn\,container1 * (Weight,\ 2/tonne)(Color,\ green),$$

and after fulfilling the final step of constructing a SR of Qs1, one is able to obtain the expression

$$Question(x1,\ Situation(e1,\ delivery2 * (Goal - place,\ x1)$$

$$(Object1,\ certn\,container1 * (Weight,\ 2/tonne)(Color,\ green)))).$$

Often used notations. We'll define in what follows some rules $P[2], \ldots, P[10]$. For $k = 1, \ldots, 10$, the rule $P[k]$ states that a certain formula b belongs to $L(B)$, a certain formula $b\,\&\,t$ belongs to the set $T^k(B)$, where $t \in Tp(S(B))$, and a certain formula z belongs to the set $Y^k(B)$. If $1 \le s \le 10$, B is any conceptual basis, then the rules $P[0], P[1], \ldots, P[s]$ determine by conjoint induction the sets of formulas

$$L(B),\ T^0(B),\ T^1(B), \ldots,\ T^s(B),\ Y^1(B), \ldots,\ Y^s(B).$$

Let's denote these sets by

$$Lnr_s(B),\ T^0(B),\ Tnr_s^1(B), \ldots,\ Tnr_s^s(B),\ Ynr_s^1(B), \ldots,\ Ynr_s^s(B)$$

and denote the family consisting of all these sets by $Globset_s(B)$.

Let $n > 1$, $Z_1, \ldots, Z_n \in Globset_s(B)$, $w_1 \in Z_1, \ldots, w_n \in Z_n$.

Then, if these relationships for the formulas $w_1, \ldots, w_n$ are the consequence of employing some rules $P[l_1], \ldots, P[l_m]$, where $m \geq 1$, we'll denote this fact by the expression of the form

$$B(l_1, \ldots, l_m) \Rightarrow w_1 \in Z_1, \ldots, w_n \in Z_n.$$

The sequence $l_1, \ldots, l_m$ may contain the repeated numbers.

In the expressions of the kind we'll often omit the symbol B in the designations of the sets $Z_1, \ldots, Z_n$; besides, we'll use the expressions $w_1, w_2 \in Z_1$, $w_3, w_4, w_5 \in Z_2$, and so on.

Example 2. It was shown above that

$$B_1(0,\ 1) \Rightarrow all\ person,\ all\,tour.group \in Lnr_1(B_1),$$

$$B_1(0,\ 1) \Rightarrow all\ person\,\&\,\{ints * dyn.phys.ob\},$$

$$all\,tour.group\,\&\,\{\{ints * dyn.phys.ob\}\} \in Tnr_1^1(B_1),$$

$$B_1(0,1) \Rightarrow all\,\&\,person\,\&\,all\ person \in Ynr_1^1(B_1),$$

$$all\,\&\,tour.group\,\&\,all\,tour.group \in Ynr_1^1(B_1).$$

where *tour.group* is the designation of the notion "a tourist group," *ints* is the sort "intelligent system."

The expression

$$B_1(0,\ 1) \Rightarrow all\,\&\,person\,\&\,all\ person \in Ynr_1^1(B_1)$$

is equivalent to the expression

$$B_1(0,1) \Rightarrow all\,\&\,person\,\&\,all\ person \in Ynr_1^1.$$

4.3 The Use of Relational Symbols and the Marking-Up of Formulas

This section introduces and illustrates the application of the rules $P[2] - P[5]$ intended for building semantic representations of NL-texts.

4.3.1 The Rules for Employing Relational Symbols

The rule $P[2]$ enables us, in particular, to construct the K-strings of the form $f(a_1, \ldots, a_n)$, where f is a designation of a function with n arguments $a_1, \ldots, a_n$.

The rule $P[3]$ is intended for building the K-strings of the form $(a_1 \equiv a_2)$, where a_1 and a_2 denote the entities characterized by the types being comparable with respect to the relation $\vdash$.

Using consecutively the rules $P[2]$ and $P[3]$, we can build the K-strings of the form $(f(a_1, \ldots, a_n) \equiv b)$, where b is the value of the function f for $a_1, \ldots, a_n$.

Let's recall that, according to the definitions given in Chap. 3, for arbitrary sort system S, the set of main types

$$Mtp(S) = Tp(S) \setminus \{[\uparrow entity], [\uparrow concept], [\uparrow object]\}.$$

Definition 4.5. Let B be any conceptual basis, $S = S(B)$. Then

- $R_1(B)$ is the set of all such $d \in X(B)$ that for each d there is such $t \in Mtp(S)$ that (a) t has no beginning "(" and (b) $tp(d)$ is the string of the form $\{t\}$;
- for arbitrary $n > 1$, $R_n(B) = \{d \in X(B) \mid$ there are such $t_1, \ldots, t_n \in Mtp(S)$ that the type $tp(d)$ is the string of the form $\{(t_1, \ldots, t_n)\}\}$;
- for arbitrary $n > 1$, $F_n(B) = F(B) \cup R_{n+1}(B)$.

If $n > 1$, the elements of $R_n(B)$ will be called *n-ary relational symbols,* and the elements of $F_n(B)$ will be called additionally *n-ary functional symbols.*

It is easy to show that for arbitrary conceptual basis B and arbitrary $k, m > 1$, it follows from $k \neq m$ that $R_k(B) \cap R_m(B) = \emptyset$.

Definition 4.6. Denote by $P[2]$ the assertion "Let $n \geq 1$, $f \in F_n(B)$,

$$tp = tp(B),\ u_1, \ldots, u_n, t \in Mtp(S(B)),$$

$$tp(f) = \{(u_1, \ldots, u_n, t)\};$$

for $j = 1, \ldots, n$, $0 \leq k[j] \leq i$, $z_j \in Mtp(S(B))$, $a_j \in L(B)$; the string $a_j \& z_j$ belong to $T^{k[j]}(B)$; if a_j doesn't belong to the set of variables $V(B)$, then $u_j \vdash z_j$ (i.e. the type z_j is a concretization of the type u_j); if $a_j \in V(B)$, then u_j and z_j are comparable with respect to the concretization relation $\vdash$. Let b be the string of the form $f(a_1, \ldots, a_n)$. Then

$$b \in L(B),\ b \& t \in T^2(B),$$

$$f \& a_1 \& \ldots \& a_n \& b \in Y^2(B)."$$

It should be recalled before the formulation of the next definition that, according to the definition of conceptual basis, the symbol $\equiv$ is an element of the primary informational universe $X(B)$ for arbitrary conceptual basis B.

Definition 4.7. Denote by $P[3]$ the assertion "Let $a_1, a_2 \in L(B)$, the types u_1, u_2 belong to the set of main types $Mtp(S(B))$, u_1 and u_2 are comparable with respect to the concretization relation $\vdash$. Let for $m = 1, 2$, $0 \leq k[m] \leq i$, the string $a_m \& u_m \in T^{k[m]}$; P be the sort "a meaning of proposition" of the conceptual basis B, and b be the string $(a_1 \equiv a_2)$. Then

$$b \in L(B),\ b\,\&\,P \in T^3(B),$$

and the string $a_1\,\&\ \equiv\ \&\,a_2\,\&\,b$ belongs to the set $Y^3(B)$."

In the rules $P[2]$ and $P[3]$, the symbol i designates an unknown integer, such that $2 \le i \le 10$. The interpretation of the symbol i is as follows:

The rules $P[0] - P[3]$ and further rules will be used together with a definition joining all these rules and having the initial phrase "Let B be an arbitrary conceptual basis, $1 \le i \le 10$." The number i will be interpreted as the maximal ordered number of a rule in the collection of the rules which will be used for constructing the formulas. For instance, if $i = 3$, we may use the rules with the numbers 0–3, but we can't employ the rules with the numbers 4–10. The parameter i in the rules for constructing the formulas enables us to define the language Lnr_{i+1} after introducing the rule $P[i+1]$ and to study the expressive possibilities of this language.

Example 1. Let B_1 be the conceptual basis defined in Chap. 2; $i = 3$;

$$b_1 = Suppliers(Firm\_Rainbow),$$

$$b_2 = Numb(Suppliers(Firm\_Rainbow)),$$

$$b_3 = (Numb(Suppliers(Firm\_Rainbow)) \equiv 12),$$

$$b_4 = Numb(all\,concept),$$

$$b_5 = Numb(all\,chemist),$$

$$b_6 = (all\,chemist \equiv S1),$$

$$b_7 = Numb(S1).$$

Then one can easily verify the validity of the following relationships (taking into account the notation introduced in the preceding section):

$$B_1(0) \Rightarrow Suppliers\,\&\,\{(org,\ \{org\})\} \in T^0,$$

$$Firm\_Rainbow\,\&\,org * space.ob * ints \in T^0,$$

$$Numb\,\&\,\{(\{[entity]\},\ nat)\} \in T^0;$$

$$B_1(0,\ 1,\ 2) \Rightarrow b_1\,\&\,\{org\} \in Tnr_3^2;$$

$$B_1(0,\ 2,\ 2) \Rightarrow b_2 \in Lnr_3,\ b_2\,\&\,nat \in Tnr_3^2,$$

$$Numb\,\&\,b_1\,\&\,b_2 \in Ynr_3^2;$$

$$B_1(0,\ 2,\ 2,\ 0,\ 3) \Rightarrow b_3 \in Lnr_3,\ b_3\,\&\,prop \in Tnr_3^3;$$

$$B_1(0,\ 1,\ 2) \Rightarrow b_4,\ b_5 \in Lnr_3,$$

$$b_4\,\&\,nat,\ b_5\,\&\,nat \in Tnr_3^2;$$

$$B_1(0,\ 1,\ 3) \Rightarrow b_6 \in Lnr_3,\ b_6\,\&\,prop \in Tnr_3^3;$$

$$S1 \in V(B_1),\ tp(S1) = \{[entity]\} \Longrightarrow$$

$$B_1(0,2) \Rightarrow b_7 \in Lnr_3,\ b_7 \,\&\, nat \in Tnr_3^2.$$

Definition 4.8. Denote by $P[4]$ the assertion "Let

$$n \geq 1,\ r \in R_n(B) \setminus F(B),\ u_1, \ldots, u_n \in Mtp(S(B)),\ tp = tp(B),$$

$tp(r)$ be the string of the form $\{(u_1, \ldots, u_n)\}$ in case $n > 1$ or the string of the form $\{u_1\}$ in case $n = 1$; for $j = 1, \ldots, n,\ 0 \leq k[j] \leq i,\ z_j \in Mtp(S(B)),\ a_j \in L(B)$;

the strings $a_j \,\&\, z_j$ belong to $T^{k[j]}(B)$; if a_j doesn't belong to the set of variables $V(B)$, then $u_j \vdash z_j$ (i.e. the type z_j is a concretization of the type u_j); if $a_j \in V(B)$, then u_j and z_j are comparable with respect to the concretization relation $\vdash$.

Let b be the string of the form $r(a_1, \ldots, a_n)$, $P = P(B)$ be the sort "a meaning of proposition" of the conceptual basis B. Then

$$b \in L(B),\ b \,\&\, P \in T^4(B),$$

$$r \,\&\, a_1 \,\&\, \ldots \,\&\, a_n \,\&\, b \in Y^4(B)."$$

Example 2. Let B_1 be the conceptual basis defined in Chapter 3; $i = 4$;

$$b_8 = Less(10000, Numb(all\ chemist)),$$

$$b_9 = Less(5000, Numb(all\ concept)),$$

$$b_{10} = Elem(P.Somov,\ all\ person),$$

$$b_{11} = Elem(Firm\_Ocean,\ Suppliers(Firm\_Rainbow)),$$

$$b_{12} = (R.Scott \equiv Director(Firm\_Ocean)),$$

$$b_{13} = Knows(N.Cope,\ (Numb(Suppliers(Firm\_Rainbow)) \equiv 12)),$$

$$b_{14} = Knows(P.Somov,\ (R.Scott \equiv Director(Firm\_Ocean))),$$

$$b_{15} = Less(10000, Numb(S1)),$$

$$b_{16} = Subset(all\ chemist,\ all\ person).$$

Taking into account the definition of the concept-object system introduced in the preceding chapter and employing the rules $P[0], \ldots, P[4]$, we have, obviously, the following relationships:

$$B_1(0,1,2,3,4) \Rightarrow b_8, \ldots, b_{16} \in Lnr_4,$$

$$b_8 \,\&\, prop, \ldots,\ b_{16} \,\&\, prop \in Tnr_4^4;$$

$$Subset \,\&\, all\ chemist \,\&\, all\ person \,\&\, Subset(all\ chemist,\ all\ person) \in Ynr_4^4.$$

Here the string *prop* is to be interpreted as the distinguished sort "a meaning of proposition," it belongs to the set of sorts $St(B_1)$.

4.3.2 The Rule for Marking Up the Formulas

The rule $P[5]$ is intended, in particular, for marking by variables in semantic representations (SRs) of NL-texts: (a) the descriptions of diverse entities mentioned in a text (physical objects, events, notions, etc.), (b) the fragments being SRs of sentences and of larger parts of texts to which a reference is given in any part of a text.

Definition 4.9. Denote by $P[5]$ the assertion "Let $a \in L(B) \setminus V(B)$, $0 \leq k \leq i$, $k \neq 5$, $t \in Mtp(S(B))$, $a \& t \in T^k(B)$; $v \in V(B)$, $u \in Mtp(S(B))$, $v \& u \in T^0(B)$, $u \vdash t$, v be not a substring of the string a. Let b be the string of the form $a : v$. Then the relationships

$$b \in L(B),\ b \& t \in T^5(B),$$

$$a \& v \& b \in Y^5(B)$$

take place."

Example 3. Consider, as before, the conceptual basis B_1 constructed in the preceding chapter. Let $i = 5$, $a_1 = b_3 = (Numb(Suppliers(Firm\_Rainbow)) \equiv 12)$, $k_1 = 3$, $t_1 = prop = P(B_1)$, that is, the informational item *prop* is the distinguished sort "a meaning of proposition." Then, obviously, $a_1 \& t_1 \in Tnr_5^3$.

Suppose that $v_1 = P1$, $z_1 = prop = P(B_1)$. Then it follows from the definition of the conceptual basis B_1 and the rule $P[0]$ that

$$v_1 \& z_1 \in T^0(B_1).$$

Besides, $z_1 \vdash t_1$ (because $z_1 = t_1$), and v_1 is not a substring of the expression a_1.

Let $b_{17} = (Numb(Suppliers(Firm\_Rainbow)) \equiv 12) : P1$. Then, according to the rule $P[5]$,

$$b_{17} \in Lnr_5(B_1),\ b_{17} \& prop \in Tnr_5^5(B_1),$$

$$a_1 \& v_1 \& b_{17} \in Ynr_5^5(B_1).$$

Let $b_{18} = Suppliers(Firm\_Rainbow) : S2$. Then it is easy to see that

$$B_1(0, 2, 5) \Rightarrow b_{18} \in Lnr_5,\ b_{18} \& \{org\} \in Tnr_5^5,$$

$$Suppliers(Firm\_Rainbow) \& S2 \& b_{18} \in Ynr_5^5.$$

In the expression b_{17}, the variable $P1$ marks up the semantic representation of the phrase T1 = "The firm 'Rainbow' has 12 suppliers." That is why if the expression b_{17} is a fragment of a long formula, then it is possible to use the mark (the variable) $P1$ to the right from the occurrence of b_{17} for repeatedly representing (if necessary) the meaning of the phrase T1 instead of the much longer semantic representation of the phrase T1.

In the expression b_{18}, the variable $S2$ marks up the set consisting of all suppliers of the firm "Rainbow."

It should be underlined that the rule $P[5]$ is very important for building semantic representations of discourses. It allows us to form SRs of discourses in such a manner that the SRs reflect the referential structure of discourses. The examples of the kind may be found in next sections.

4.4 The Use of Logical Connectives NOT, AND, OR

In comparison with the first-order predicate logic, the rules $P[6]$ and $P[7]$ in combination with the other rules allow for more complete modeling (on the level of semantic representations of NL-texts) of the manners of employing the logical connectives "not," "and," "or" in sentences and discourses in English, Russian, German, and many other languages. In particular, the existence of sentences of the kinds "This medicine was produced not in UK," "Professor Cope defended his Ph.D. dissertation not in the Stanford University," "This patent has been used in Austria, Hungary, The Netherlands, and France," has constructively been taken into account.

With this aim, first, it is permitted to join the connective $\neg$ not only to the expressions designating statements but also to the denotations of things, events, and notions. Second, it is permitted to use the connectives $\wedge$ (conjunction, logical "and") and $\vee$ (disjunction, logical "or") not only for joining the semantic representations of the statements but also for joining the denotations of things, events, and notions.

The rule $P[6]$ describes the use of the connective $\neg$ ("not").

Definition 4.10. Denote by $P[6]$ the assertion "Let $a \in L(B)$, $t \in Mtp(S(B))$, $0 \leq k \leq i$, k be not in the set $\{2, 5, 10\}$, $a \,\&\, t \in T^k(B)$, b be the string of the form $\neg\, a$. Then $b \in L(B)$, $b \,\&\, t \in T^6(B)$, the string of the form $\neg \,\&\, a \,\&\, b$ belongs to $Y^6(B)$."

In this assertion, the expression $\{2, 5, 10\}$ designates the set consisting of several forbidden values for k. This means the following: if an l-formula a is constructed step by step in any way with the help of some rules from the list $P[0], P[1], \ldots, P[i]$, then the rule $P[2]$ or $P[5]$ or $P[10]$ can't be applied at the last step of the inference.

Example 1. If B_1 is conceptual basis defined in the preceding chapter, $i = 6$, then one can easily verify that

$$B_1(0, 6) \Rightarrow \neg\, chemist \in Lnr_6,$$

$$chemist \,\&\, \uparrow ints * dyn.phys.ob \in Tnr_6^6;$$

$$B_1(0, 6, 4, 4) \Rightarrow Knows(P.Somov,\ Is1(N.Cope,\ \neg\, chemist)) \in Lnr_6,$$

$$B_1(0, 4, 4, 6) \Rightarrow \neg\, Knows(P.Somov,\ Is1(N.Cope,\ biologist)) \in Lnr_6.$$

We'll interpret the built l-formulas as possible semantic representations of the NL-expressions "not a chemist," "P. Somov knows that N. Cope is not a chemist," and "P. Somov doesn't know that N. Cope is a biologist."

The rule $P[7]$ describes the manners to use the connectives $\wedge$ (conjunction, logical "and") and $\vee$ (disjunction, logical "or") while constructing semantic representations of NL-texts. This rule allows for building, in particular, the l-formulas of the form $(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$ and of the form $(a_1 \vee a_2 \vee \ldots \vee a_n)$.

For instance, the rule $P[7]$ jointly with some other rules enables us to construct the l-formulas

$$(chemist \vee biologist),\ (mathematician \wedge painter),$$

$$(First.name(x1, Pavel) \wedge Surname(x1, Somov)$$
$$\wedge Qualification(x1, chemist)).$$

Definition 4.11. Denote by $P[7]$ the assertion "Let

$$n > 1,\ t \in Mtp(S(B)),\ \text{for}\, m = 1, \ldots, n,$$

$$0 \leq k[m] \leq i,\ a_m \in L(B),\ a_m \,\&\, t \in T^{k[m]}(B);$$

$s \in \{\wedge, \vee\}$, b be the string of the form

$$(a_1\, s\, a_2\, s \ldots s\, a_n).$$

Then the relationships

$$b \in L(B),\ b \,\&\, t \in T^7(B),$$
$$s \,\&\, a_1 \,\&\, \ldots \,\&\, a_n \,\&\, b \in Y^7(B)$$

take place."

Comment to the rule $P[7]$. According to this rule, all expressions joined during one step by a logical connective, are to be associated with the same type. Since this type t can be different from the distinguished sort "meaning of proposition," it is possible to employ the rule $P[7]$ for joining with the help of binary logical connectives not only semantic representations of statements but also the designations of various objects, simple and compound designations of the notions and of the goals of intelligent systems.

Example 2. Suppose that B_1 is the conceptual basis defined in the preceding chapter, $i = 6$. Let

$$b_1 = (R.Scott \wedge N.Cope),$$

$$b_2 = (chemist \vee biologist),$$

$$b_3 = ((Numb(Friends(J.Price)) \equiv 3) : P1$$
$$\wedge Knows(P.Somov, now, P1) \wedge$$
$$\neg Knows(P.Somov, now, Is1(J.Price, (chemist \vee biologist)))),$$

$$b_4 = Knows(P.Somov, now, Elem((R.Scott \wedge N.Cope), Friends(J.Price))),$$

$$b_5 = (Elem(R.Scott, Friends(N.Cope) : S3)$$
$$\wedge \neg Elem(P.Somov, S3)).$$

It is not difficult to show that

$$B_1(0, 7) \Rightarrow b_1,\ b_2 \in Lnr_7,$$

$$b_1 \,\&\, ints * dyn.phys.ob \in Tnr_7^7;$$

$$b_2 \,\&\, \uparrow ints * dyn.phys.ob \in Tnr_7^7;$$

$$B_1(0, 2, 2, 3, 5, 4, 7, 4, 4, 6, 7) \Rightarrow b_3 \in Lnr_7,$$

$$B_1(0, 7, 2, 4, 4) \Rightarrow b_4 \in Lnr_7,\ b_4 \,\&\, prop \in Tnr_7^4;$$

$$B_1(0, 2, 5, 4, 4, 6, 7) \Rightarrow b_5 \in Lnr_7,$$

$$b_5 \,\&\, prop \in Tnr_7^7.$$

4.5 Building Compound Designations of Notions and Objects

It will be shown below how to build compound representations of notions and, if necessary, to transform these representations of notions into the compound representations of objects (things, events, etc.), applying once the rule $P[1]$ and (it is optional) once the rule $P[5]$.

4.5.1 Compound Designations of Notions

Let's consider the rule $P[8]$ intended for constructing compound representations of notions (concepts) such as

$$text.book * (Field1,\ biology),$$

$$city * (Country,\ France),$$

$$concept * (Name.of.concept,\ \prime molecule \prime),$$

$$tourist.group * (Number.of.persons,\ 12)$$

$$(Qualitative\_composition,\ (chemist \vee biologist)).$$

Together with the rule $P[l]$ and other rules, it will enable us to build compound designations of things and sets of things in the form *qtr descr*, where *qtr* is an intensional quantifier, *descr* is a compound designation of a notion formed with the help of the rule $P[8]$ at the last step of the inference.

For instance, in this way the following formulas can be built:

$$all\ person * (Age,\ 18/year),$$

$$certain\ person * (Age,\ 18/year),$$

$$certain\ tourist.group * (Number\_of\_persons,\ 12).$$

It should be recalled that for arbitrary conceptual basis B,

$$Tconc(B) = \{t \in Tp(S(B)) \mid t \text{ has the beginning } \uparrow\} \cup Spectp,$$

where

$$Spectp = \{[\uparrow entity], [\uparrow concept], [\uparrow object]\}.$$

Each element cpt of the primary universe $X(B)$ such that $tp(cpt) \in Tconc(B)$ is interpreted as a designation of a notion (a concept).

The set $R_2(B)$ consists of binary relational symbols (some of them may correspond to functions with one argument); $F(B)$ is the set of functional symbols. The element $ref = ref(B)$ from $X(B)$ is called the referential quantifier and is interpreted as a semantic item corresponding to the meaning of the expression "a certain" ("a certain book," etc.). $P(B)$ is the distinguished sort "a meaning of proposition" of the conceptual basis B.

Definition 4.12. Denote by $P[8]$ the assertion "Let

$$cpt \in X(B), tp = tp(B), t = tp(cpt), t \in Tconc(B), P = P(B), ref = ref(B).$$

Let $n \geq 1$, for $m = 1,\ldots,n$, $r_m \in R_2(B)$, c_m be the string of the form $ref\,cpt$, and d_m, $h_m \in L(B)$. If $r_m \in R_2(B) \cup F(B)$, let h_m be the string of the form $(r_m(c_m) \equiv d_m)$ and $h_m \,\&\, P \in T^3(B)$; if $r_m \in R_2(B) \setminus F(B)$, let h_m be the string of the form $r_m(c_m, d_m)$ and $h_m \,\&\, P \in T^4(B)$.

Let b be the string of the form

$$cpt * (r_1, d_1) \ldots (r_n, d_n).$$

Then the relationships

$$b \in L(B),\; b \,\&\, t \in T^8(B),$$

$$cpt \,\&\, h_1 \,\&\, \ldots \,\&\, h_n \,\&\, b \in Y^8(B).$$

take place."

Example 1. Suppose that B_1 is the conceptual basis defined in the final part of the preceding chapter, $i = 8$. Then consider a possible way of constructing the formula b_1 (defined below) corresponding to the notion "a tourist group consisting of 12 persons." Let

$$cpt = tour.group,\; t = tp(cpt) = \uparrow \{ints * dyn.phys.ob\},$$

$$P = prop,\; ref = certn,\; c_1 = 1,\; r_1 = Numb,$$

$$c_1 = ref\,cpt = certn\,tour.group,\; d_1 = 12,$$

$$h_1 = (r_1(c_1) \equiv d1) = (Numb(certn\,tour.group) \equiv 12).$$

Then

$$B_1(0, 1, 2, 3) \Rightarrow h_1 \in Lnr_3(B_1),$$

$$h_1 \,\&\, prop \in Tnr_8^3(B_1).$$

Let $b_1 = cpt * (r_1, d_1) = tour.group * (Numb, 12)$. Then it follows from the rule $P[8]$ that

$$b_1 \in Lnr_8(B_1),$$

$$b_1 \,\&\, \uparrow \{ints * dyn.phys.ob\} \in Tnr_8^8.$$

4.5.2 Compound Designations of Objects

Together with the rule $P[1]$ and other rules, the rule $P[8]$ allows for building compound designations of things, events, sets of things, and sets of events (all these entities are considered in this book as the particular kinds of objects). The compound designations of objects are constructed in the form

$$qtr\, cpt * (r_1, d_1) \ldots (r_n, d_n)$$

or in the form

$$qtr\, cpt * (r_1, d_1) \ldots (r_n, d_n) \; : \; v,$$

where qtr is an intensional quantifier, $cpt * (r_1, d_1) \ldots (r_n, d_n)$ is a compound designation of a notion formed with the help of the rule $P[8]$ at the last step of the inference, and v is a variable interpreted as an individual mark of the considered object.

For instance, in this way the following formulas can be constructed:

$$certain\, person * (Age, 18/year),$$

$$certain\, person * (Age, 18/year) \; : \; x5,$$

$$certain\, tourist.group * (Number.of.persons, 12),$$

$$certain\, tourist.group * (Number.of.persons, 12) \; : \; y3,$$

$$all\, person * (Age, 18/year),$$

$$all\, person * (Age, 18/year) \; : \; S1.$$

Example 2. Let's proceed from the same assumptions concerning the conceptual basis B_1 and the integer i as in Example 1. Then consider a possible way of building the formula b_2 (defined below) denoting a person characterized by the expression "a certain biologist from a certain tourist group consisting of 12 persons." Let

$$b_2 = certn\, biologist * (Elem, certn\, tour.group * (Numb, 12)).$$

Then

$$B_1(0, 1, 2, 3, 8, 1, 1, 4, 8, 1) \Rightarrow$$

$$b_2 \in Lnr_8(B_1), \; b_2 \,\&\, ints * dyn.phys.ob \in Tnr_8^1.$$

Example 3. Suppose that B_1 is the conceptual basis defined in the final part of the preceding chapter, $i = 8$. Then let's consider a way of constructing a possible semantic representation of the phrase "R. Scott has included N. Cope into a tourist group consisting of 12 persons." Let the variable $x1$ denote a moment of time, the variable $S3$ denote a concrete tourist group, and

$$b_3 = (Include1(R.Scott, N.Cope, x1, certn\,tour.group*$$

$$(Numb, 12) : S3 \wedge Less(x1, \#now\#)).$$

Then

$$B_1(0, 1, 2, 3, 8, 1, 4, 5, 0, 4, 7) \Rightarrow$$

$$b_3 \in Lnr_8(B_1),\ b_3 \,\&\, prop \in Tnr_8^7.$$

4.6 Final Rules

This section introduces the rules $P[9]$ and $P[10]$ intended respectively for (a) the employment in the formulas of the existential and universal quantifiers, (b) constructing the representations of finite sequences.

4.6.1 The Use of Existential and Universal Quantifiers

The rule $P[9]$ describes how to join existential quantifier $\exists$ and universal quantifier $\forall$ to the semantic representations of statements (assertions, propositions). The distinctions from the manner of using these quantifiers in the first-order predicate logics are as follows: (a) the sphere of acting of quantifiers is explicitly restricted; (b) the variables used together with the quantifiers can denote not only the things, the numbers, etc., but also the sets of various entities.

For example, we'll be able to build a semantic representation of the sentence "For each country in Europe, there is a city with the number of inhabitants exceeding 3000" in the form

$$\forall x1\,(country * (Location1, Europe))\,\exists x2(city)$$

$$((Location1(x2, x1) \wedge Less(3000, Numb(Inhabitants(x2)))).$$

Here the expressions country $*(Location1, Europe)$ and $city$ restrict the domain where the variables $x1$ and $x2$ can take values respectively.

Definition 4.13. Let's denote by $P[9]$ the assertion "Let $qex \in \{\exists, \forall\}$,

$$A \in L(B) \setminus V(B),\ P = P(B),\ k \in \{3, 4, 6, 7, 9\},$$

$$a \,\&\, P \in T^k(B),\ var \in V(B),\ tp = tp(B),$$

$tp(var) = [entity]$ is the basic type "entity," the string A includes the symbol var, $m \in \{0, 8\}$,

$$concept\_denot \in L(B) \setminus V(B),\ u \in Tc(B),$$

where $Tc(B)$ is the set of all types from the set $Tp(S(B))$ having the beginning $\uparrow$; the string $concept\_denot \,\&\, u$ belongs to the set $T^m(B)$.

Besides, let the string A don't include the substrings of the forms : var, $\exists var$, $\forall var$ and A don't have the ending of the form : z, where z is an arbitrary variable from $V(B)$, b be the string of the form

$$qex\ var\,(concept\_denot)\,A.$$

Then the relations

$$b \in L(B),\ b \,\&\, t \in T^9(B),$$

$$qex \,\&\, var \,\&\, \ concept\_denot \,\&\, A \,\&\, b \in Y^9(B)$$

take place."

Example 1. Let's construct a possible SR of the phrase T1 = "There are such a moment $x1$ and a tourist group $S3$ consisting of 12 persons that R. Scott included N.Cope into the group $S3$ at the moment $x1$." Let B_1 be the conceptual basis determined in the preceding chapter, $i = 9$, and

$$b_4 = \exists x1\,(mom)\,\exists S3(tour.group * (Numb,\ 12))$$

$$(Include1(R.Scott,\ N.Cope,\ x1,\ S3) \wedge Before(x1,\ \#now\#)).$$

Then it is easy to show that

$$B_1(0,\ 4,\ 4,\ 7,\ 0,\ 1,\ 2,\ 3,\ 8,\ 9,\ 9) \Rightarrow$$

$$b_4 \in Lnr_9(B_1),\ b_4 \,\&\, prop \in Tnr_9^9.$$

The formula b_4 is to be interpreted as a possible SR of T1.

Example 2. Let $i = 9$ and there is such conceptual basis B that the following relationships take place:

$$space.ob,\ nat.number,\ dyn.phys.ob,\ ints \in St(B),$$

$$city,\ country,\ Europe,\ Location1,\ Number\_of\_elem,$$

$$Inhabitants,\ Less \in X(B),$$

$$tp(city) = tp(country) = \uparrow space.ob,$$

$$tp(Europe) = space.ob,$$

$$tp(Location1) = \{(space.ob,\ space.ob)\},$$

$$tp(Number\_of\_elem) = \{(\{[entity]\},\ nat.number)\},$$

$$tp(Inhabitants) = \{(space.ob,\ \{dyn.phys.ob * ints\})\},$$

$$tp(Less) = \{(nat.number,\ nat.number)\},$$

$$Number\_of\_elem,\ Inhabitants \in F(B),$$

$$x1, x2 \in V(B),\ tp(x1) = tp(x2) = [entity],$$

$$3000 \in X(B),\ tp(3000) = nat.number.$$

$$prop = P(B).$$

The listed informational units are interpreted as follows:

$$space.ob,\ nat.number,\ dyn.phys.ob,\ ints$$

are the sorts "space object," "natural number," "dynamic physical object," "intelligent system"; *city*, *country* are the designations of the notions "a city" and "a country"; *Europe* is a designation of the world part Europe; *Location*1 is a designation of a binary relation between space objects; *Number_of_elem* is a designation of the function "Number of elements of a set"; *Inhabitants* is a designation of the function associating with every locality (a village, a city, a country, etc.) the set consisting of all inhabitants of this locality; *Less* is a designation of a binary relation on the set of natural numbers.

Let $qex_1 = \exists$, $var_1 = x2$, $concept\_denot_1 = city$,

$$A_1 = (Location1(x2, x1) \wedge Less(3000, Number\_of\_elem(Inhabitants(x2)))),$$

$$b_1 = qex_1\, var_1\, (concept\_denot_1)A_1.$$

Then it is not difficult to see that the following relationship takes place:

$$B(0, 2, 2, 3, 4, 4, 7, 9) \Rightarrow$$

$$b_1 \in Lnr_9(B),\ b_1 \,\&\, prop \in Tnr_9^9.$$

Let $qex_2 = \forall$, $var_2 = x1$,

$$concept\_denot_2 = country * (Location1, Europe),$$

$$A_2 = b_1,\ b_2 = qex_2\, var_2\, (concept\_denot_2)A_2.$$

Then it is easy to verify that

$$B(0, 2, 2, 3, 4, 4, 7, 9, 0, 1, 4, 8, 9) \Rightarrow$$

$$b_2 \in Lnr_9(B),\ b_2 \,\&\, prop \in Tnr_9^9.$$

4.6.2 The Representations of Finite Sequences

The rule $P[10]$ is intended for building the representations of finite sequences consisting of n elements, where $n > 1$, in the form $\langle a_1, a_2, \ldots, a_n \rangle$; such sequences are usually called in mathematics the n-tuples.

Definition 4.14. Denote by $P[10]$ the assertion "Let $n > 1$, for $m = 1, \ldots, n$, the following relationships take place:

$$a_m \in L(B),\ u_m \in Mtp(S(B)),$$

$$0 \leq k[m] \leq 10,\ a_m \,\&\, u_m \in T^{k[m]}.$$

Let t be the string of the form $(u_1, u_2, \ldots, u_n)$, and b be the string of the form

$$\langle a_1, a_2, \ldots, a_n \rangle.$$

Then the relationships

$$b \in L(B),\ b \,\&\, t \in T^{10}(B),$$

$$a_1 \,\&\, a_2 \,\&\, \ldots \,\&\, a_n \,\&\, b \in Y^{10}(B)$$

take place."

Example 3. Let B_1 be the conceptual basis defined in Chapter 3, $i = 10$, and b_3 be the string of the form

$$(Elem(x3, S1) \equiv ((x3 \equiv \langle certn\,real : x1,\ certn\,real : x2 \rangle) \wedge$$

$$(Less(x1, x2) \vee (x1 \equiv x2)))),$$

where the string b_3 is to be interpreted as a possible formal definition of the binary relation "Less or equal" on the set of real numbers. Then one can easily show that

$$B(0, 4, 1, 5, 1, 5, 4, 10, 3, 7, 7, 3) \Rightarrow$$

$$b_3 \in Lnr_{10}(B),\ b_3 \,\&\, prop \in Tnr^3_{10}.$$

4.6.3 A Summing-Up Information about the Rules P[0]–P[10]

The total volume of the definitions of the rules $P[0] - P[10]$ and of the examples illustrating these rules is rather big. For constructing semantic representations not only of discourses but also of a major part of separate sentences, one is to use a considerable part of these rules, besides, in numerous combinations.

Taking this into account, it seems to be reasonable to give a concise, non detailed characteristic of each rule from the list $P[0] - P[10]$. The information given below will make easier the employment of these rules in the course of constructing SRs of NL-texts and representing the pieces of knowledge about the world.

Very shortly, the principal results (the kinds of constructed formulas) of applying the rules $P[0] - P[10]$ are as follows:

- $P[0]$: An initial stock of l-formulas and t-formulas determined by (a) the primary informational universe $X(B)$, (b) the set of variables $V(B)$, and (c) the mapping tp giving the types of the elements from these sets.

- $P[1]$: l-formulas of the kind $qtr\, cpt$ or $qtr\, cpt * (r_1, d_1) \dots (r_n, d_n)$, where qtr is an intensional quantifier, cpt is a simple (i.e., non structured) designation of a notion (a concept), $n \geq 1$, $r_1, \dots, r_n$ are the designations of the characteristics of the entities (formally, these designations are unary functional symbols or the names of binary relations).
- $P[2]$: l-formulas of the kind $f(a_1, \dots, a_n)$, where f is a functional symbol, $n \geq 1$, $a_1, \dots, a_n$ are the designations of the arguments of the function with the name f; t-formulas of the kind $f(a_1 \dots a_n) \,\&\, t$, where t is the type of the value of the function f for the arguments $a_1, \dots, a_n$.
- $P[3]$: l-formulas of the kind $(a_1 \equiv a_2)$ and t-formulas of the kind $(a_1 \equiv a_2) \,\&\, P$, where P is the distinguished sort "a meaning of proposition."
- $P[4]$: l-formulas of the kind $rel\,(a_1, \dots, a_n)$, where r is a relational symbol, $n \geq 1$, $a_1, \dots, a_n$ are the designations of the attributes of the relation with the name rel; t-formulas of the kind $rel\,(a_1, \dots, a_n) \,\&\, P$, where P is the distinguished sort "a meaning of proposition."
- $P[5]$: l-formulas of the kind $form\, :\, v$, where $form$ is l-formula, v is a variable from $V(B)$ being a mark of the formula $form$.
- $P[6]$: Proceeding from the l-formula $form$, in particular, the l-formula of the kind $\neg form$ is constructed.
- $P[7]$: Using, as the operands, a logical connective $s \in \{\vee, \wedge\}$ and some l-formulas $a_1, \dots, a_n$, where $n > 1$, one obtains, in particular, the l-formula

$$(a_1\, s\, a_2\, s\, \dots\, s\, a_n).$$

- $P[8]$: The operands of this rule are (a) an l-formula cpt from the primary informational universe $X(B)$ interpreted as a simple (non structured) designation of a notion (a concept), (b) the characteristics $r_1, \dots, r_n$, where $n > 1$, of the entities qualified by the notion cpt, (c) the l-formulas $d_1, \dots, d_n$; for $k = 1, \dots, n$, if r_k is the name of a function with one argument, d_k is the value of this function for a certain entity qualified by the notion cpt; if r_k is the name of a binary relation, d_k designates the second attribute of this relation, where the first attribute is a certain entity qualified by the notion cpt. These operands are used for constructing (a) the l-formula
$$cpt * (r_1, d_1) \dots (r_n, d_n)$$
and (b) the t-formula
$$cpt * (r_1, d_1) \dots (r_n, d_n) \,\&\, t,$$
where t is a type from the set $Tp(S(B))$; such l-formula and t-formula are interpreted as the compound designations of the notions (concepts). For instance, there is such a conceptual basis B that the l-formula $city * (Country, France)$ and the t-formula $city * (Country, France) \,\&\, space.object$ can be constructed.
- $P[9]$: The l-formulas of the kind
$$qex\, var\, (concept\_denot)\, A$$
and t-formulas of the kind

$$qex\,var\,(concept\_denot)\,A\,\&\,P$$

are constructed, where *qex* is either the existential quantifier $\exists$ or the universal quantifier $\forall$, *var* is a variable from $V(B)$, *concept_denot* is a simple (non structured) or compound designation of a notion (a concept), A is an l-formula reflecting the semantic content of a statement, P is the distinguished sort "a meaning of proposition."

- $P[10]$: The l-formulas of the kind $\langle a_1, \ldots, a_n \rangle$, where $n > 1$, are constructed; such formulas are interpreted as the designations of finite sequences containing n elements, or n-tuples.

4.7 SK-Languages: Mathematical Investigation of Their Properties

Let's remember the denotations introduced in the first section of this chapter. Suppose that B is an arbitrary conceptual basis, and *Specsymbols* is the set consisting of the symbols $\prime,\prime$ (comma), $\prime(\prime, \prime)\prime, \prime:\prime, \prime*\prime, \prime\langle\prime, \prime\rangle\prime$. Then

$$D(B) = X(B) \cup V(B) \cup Specsymbols,$$

$$Ds(B) = D(B) \cup \{\prime\&\prime\},$$

$D^+(B)$ and $Ds^+(B)$ are the sets of all non empty finite sequences of the elements from $D(B)$ and $Ds(B)$ respectively.

Definition 4.15. Let B be an arbitrary conceptual basis, $1 \leq i \leq 10$, and the sets of strings

$$L(B) \subset D^+(B),$$

$$T^0(B), T^1(B), \ldots, T^i(B) \subset Ds^+(B),$$

$$Y^1(B), \ldots, Y^i(B) \subset Ds^+(B)$$

are the least sets jointly determined by the rules $P[0], P[1], \ldots, P[i]$. Then denote these sets, respectively, by

$$Lnr_i(B),\ T^0(B),\ Tnr_i^1(B), \ldots, Tnr_i^i(B),$$

$$Ynr_i^1(B), \ldots, Ynr_i^i(B)$$

and denote the family (that is, the set) consisting of all listed sets of strings by $Globset_i(B)$.

Besides, let the following relationships take place:

$$T_i(B) = T^0(B) \cup Tnr_i^1(B) \cup \ldots \cup Tnr_i^i(B); \tag{4.1}$$

$$Y_i(B) = Ynr_i^1(B) \cup \ldots \cup Ynr_i^i(B); \tag{4.2}$$

$$Form_i(B) = Lnr_i(B) \cup T_i(B) \cup Y_i(B). \tag{4.3}$$

Definition 4.16. If B is an arbitrary conceptual basis, then:

$$Ls(B) = Lnr_{10}(B), \tag{4.4}$$

$$Ts(B) = T_{10}(B), \tag{4.5}$$

$$Ys(B) = Y_{10}(B), \tag{4.6}$$

$$Forms(B) = Form_{10}(B), \tag{4.7}$$

$$Ks(B) = (B, Rules), \tag{4.8}$$

where *Rules* is the set consisting of the rules $P[0], P[1], \ldots, P[10]$.

The ordered pair $Ks(B)$ is called *the K-calculus (knowledge calculus) in the conceptual basis B*; the elements of the set $Forms(B)$ are called *the formulas inferred in the stationary conceptual basis B*. The formulas from the sets $Ls(B)$, $Ts(B)$, and $Ys(B)$ are called respectively l-formulas, t-formulas, and y-formulas. The set of l-formulas $Ls(B)$ is called *the standard knowledge language (or SK-language, standard K-language) in the conceptual basis B*.

The set $Ls(B)$ is considered as the main subclass of formulas generated by the collection of the rules $P[0], \ldots, P[10]$. The formulas from this set are intended for describing structured meanings (or semantic content) of NL-texts.

Theorem 4.1. *If B is an arbitrary conceptual basis, then:*
(a) the set $Lnr_0(B) \neq \emptyset$;
(b) for $m = 1, \ldots, 10$, $Lnr_{m-1}(B) \subseteq Lnr_m(B)$.

Proof. (a) For arbitrary conceptual basis B, $X(B)$ includes the non empty set $St(B)$; this follows from the definitions of a sort system and a conceptual basis. According to the rule $P[0]$, $St(B) \subset Lnr_0(B)$. Therefore, $Lnr_0(B) \neq \emptyset$. (b) The structure of the first definition in this section and the structure of the statements $P[0], P[l], \ldots, P[10]$ show that the addition of the rule $P[m]$, where $1 \leq m \leq 10$, to the list $P[0], \ldots, P[m-1]$ either expands the set of l-formulas or doesn't change it (if the set of functional symbols $F(B)$ is empty, $Lnr_1(B) = Lnr_2(B)$). The proof is complete.

Taking into account this theorem, it is easy to see that previous sections of this chapter provide numerous examples of l-formulas, t-formulas, and y-formulas inferred in a stationary conceptual basis and, as a consequence, numerous examples of the expressions of SK-languages.

Theorem 4.2. *If B is any conceptual basis, then:*

$$Ls(B),\ Ts(B),\ Ys(B) \neq \emptyset.$$

Proof. Let B be a conceptual basis. Then $Ls(B) \neq \emptyset$ according to the Theorem 4.1 and the relationship (4.4). It follows from the definition of a concept-object system that the set of variables $V(B)$ includes a countable subset of such variables *var* that

$tp(var) = [entity]$. Let v_1 and v_2 be two variables from this subset, b be the string of the form $(v_1 \equiv v_2)$, c be the string of the form

$$v_1 \,\&\, \equiv \,\&\, v_2 \,\&\, (v_1 \equiv v_2),$$

$P = P(B)$ be the sort "a meaning of proposition" of B. Then, according to the rules $P[0]$ and $P[3]$,

$$b \,\&\, P \in Tnr_i^3(B),\ c \in Ynr_i^3(B)$$

for each $i = 3, \ldots, 10$. Therefore, with respect to the relationships (4.1), (4.2), (4.5), and (4.6), the sets of formulas $Ts(B)$ and $Ys(B)$ are non empty.

Theorem 4.3. *If B is an arbitrary conceptual basis, then:*

(a) If $w \in Ts(B)$, then w is a string of the form $a \,\&\, t$, where $a \in Ls(B)$, $t \in Tp(S(B))$, and such a representation depending on w is unique for each w;

(b) if $y \in Ys(B)$, then there are such $n > 1$ and $a_1, \ldots, a_n, b \in Ls(B)$ that y is the string of the form

$$a_1 \,\&\, \ldots \,\&\, a_n \,\&\, b;$$

besides, such representation depending on y is unique for each y.

Proof. The structure of the rules $P[0], \ldots, P[10]$ and the definition of a conceptual basis immediately imply the truth of this theorem.

Theorem 4.4. *If B is a conceptual basis, $d \in X(B) \cup V(B)$, then there are no such integer k, where $1 \leq k \leq 10$, $n > 1$, and the strings $a_1, \ldots, a_n \in Ls(B)$ that*

$$a_1 \,\&\, \ldots \,\&\, a_n \,\&\, d \in Ynr_{10}^k(B).$$

Interpretation. If d is an element of the primary informational universe $X(B)$ or a variable from $V(B)$, then it is impossible to obtain d with the help of any operations determined by the rules $P[1] - P[10]$.

Proof. Assume that there are such $k \in \{1, \ldots, 10\}$, $n > 1$, $a_1, \ldots, a_n \in Ls(B)$ that the required relationship takes place. For arbitrary $m \in \{1, \ldots, 10\}$, the set $Ynr_{10}^m(B)$ may include a string $a_1 \,\&\, \ldots \,\&\, a_n \,\&\, d$, where d contains no occurrences of the symbol &, only in case d is obtained from the elements $a_1, \ldots, a_n$ by means of applying one time the rule $P[m]$. Then it follows from the structure of the rules $P[l], \ldots, P[10]$ that d must contain at least two symbols. But we consider the elements of the set $X(B) \cup V(B)$ as symbols. Therefore, we get a contradiction, since we assume that $d \in X(B) \cup V(B)$. The proof is complete.

Theorem 4.5. *Let B be an arbitrary conceptual basis,*

$$z \in Ls(B) \setminus (X(B) \cup V(B)).$$

Then there is one and only one such $n+3$-tuple $(k, n, y_0, y_1, \ldots, y_n)$, where $1 \leq k \leq 10$, $n \geq 1$, $y_0, y_1, \ldots, y_n \in Ls(B)$ that

$$y_0 \,\&\, y_1 \,\&\, \ldots \,\&\, y_n \,\&\, z \in Ynr_{10}^k(B).$$

Interpretation. If an l-formula z doesn't belong to the union of the sets $X(B)$ and $V(B)$, then there exists the unique rule $P[k]$, where $1 \leq k \leq 10$, and the unique finite sequence of l-formulas $y_0, y_1, \ldots, y_n$ that the string z is constructed from the "blocks" $y_0, y_1, \ldots, y_n$ by means of applying just one time the rule $P[k]$.

The truth of this theorem can be proved with the help of two lemmas. In order to formulate these lemmas, we need

Definition 4.17. Let B be an arbitrary conceptual basis, $n \geq 1$, for $i = 1, \ldots, n$, $c_i \in D(B)$, $s = c_1 \ldots c_n$, $1 \leq k \leq n$. Then let the expressions $lt_1(s, k)$ and $lt_2(s, k)$ denote the number of the occurrences of the symbol $'('$ and symbol $'\langle'$ respectively in the substring $c_1 \ldots c_k$ of the string $s = c_1 \ldots c_n$.

Let the expressions $rt_1(s, k)$ and $rt_2(s, k)$ designate the number of the occurrences of the symbol $')'$ and the symbol $'\rangle'$ in the substring $c_1 \ldots, c_k$ of the string s. If the substring $c_1, \ldots, c_k$ doesn't include the symbol $'('$ or the symbol $'\langle'$, then let respectively

$$lt_1(s, k) = 0,\ lt_2(s, k) = 0\,,$$

$$rt_1(s, k) = 0,\ rt_2(s, k) = 0.$$

Lemma 1. Let B be an arbitrary conceptual basis, $y \in Ls(B)$, $n \geq 1$, for $i = 1, \ldots, n$, $c_i \in D(B)$, $y = c_1 \ldots c_n$. Then

(a) if $n > 1$, then for every $k = 1, \ldots, n-1$ and every $m = 1, 2$,

$$(a) \quad lt_m(y, k) \geq rt_m(y, k);$$

$$(b) \quad lt_m(y, n) = rt_m(y, n).$$

Lemma 2. Let B be an arbitrary conceptual basis, $y \in Ls(B)$, $n > 1$, $y = c_1 \ldots c_n$, where for $i = 1, \ldots, n$, $c_i \in D(B)$, the string y include the comma or any of the symbols $\equiv$, $\wedge$, $\vee$, and k be such arbitrary natural number that $1 < k < n$. Then

(a) if c_k is one of the symbols $\equiv$, $\wedge$, $\vee$, then

$$lt_1(y, k) > rt_1(y, k) \geq 0;$$

(b) if c_k is the comma , then at least one of the following relationships takes place:

$$lt_1(y, k) > rt_1(y, k) \geq 0,$$

$$lt_2(y, k) > rt_2(y, k) \geq 0.$$

The proofs of Lemma 1, Lemma 2, and Theorem 4.5 can be found in the Appendix to this monograph.

Definition 4.18. Let B be an arbitrary conceptual basis, $z \in Ls(B) \setminus (X(B) \cup V(B))$, and there is such $n+3$-tuple $(k, n, y_0, y_1, \ldots, y_n)$, where $1 \leq k \leq 10$, $n \geq 1$, $y_0, y_1, \ldots, y_n \in Ls(B)$ that

$$y_0 \,\&\, y_1 \,\&\, \ldots \,\&\, y_n \,\&\, z \in Ynr^k_{10}(B).$$

Then the $n+3$-tuple $(k, n, y_0, y_1, \ldots, y_n)$ will be called *a form-establishing sequence of the string z.*

With respect to this definition, the Theorem 4.5 states that for arbitrary conceptual basis B, each string from $Ls(B) \setminus (X(B) \cup V(B))$ has the unique form-establishing sequence.

Theorem 4.6. *Let B be an arbitrary conceptual basis, $z \in Ls(B)$. Then there is one and only one such type $t \in Tp(S(B))$ that $z\&t \in Ts(B)$.*

Proof. Let's consider two possible cases.

Case 1.

Let B be an arbitrary conceptual basis, $z \in X(B) \cup V(B), t \in Tp(S(B)), tp(z) = t$. Then the rule $P[0]$ implies that $z\&t \in Ts(B)$.

Assume that w is such type from $Tp(S(B))$ that $z\&w \in Ts(B)$. The analysis of the rules $P[0] - P[10]$ shows that this relationship can follow only from the rule $P[0]$. But in this case, w is unambiguously determined by this rule, therefore, w coincides with t.

Case 2.

Let B be an arbitrary conceptual basis, $z \in Ls(B) \setminus (X(B) \cup V(B))$. In accordance with Theorem 4.5, there are such integer k, n where $1 \leq k \leq 10$, $n \geq 1$, and such $y_0, y_1, \ldots, y_n \in Ls(B)$ that the string z is constructed from these elements by applying one time the rule $P[k]$. That is why there is such type $t \in Tp(S(B))$ that $z\&t \in Tnr^k_{10}(B)$ and, as a consequence, $z\&t \in Ts(B)$.

It follows from Theorem 4.5 that z unambiguously determines k, n,

$y_0, y_1, \ldots, y_n$ of the kind. But in this situation, the $n+3$-tuple $(k, n, y_0, y_1, \ldots, y_n)$ unambiguously determines such type u that $z\&u \in Tnr^k_{10}(B)$. Therefore, the type u coincides with the type t.

Taking this into account, Theorem 4.5 states that every string of the SK-language $Ls(B)$, where B is an arbitrary conceptual basis, can be associated with the only type t from $Tp(S(B))$.

Definition 4.19. Let B be an arbitrary conceptual basis, $z \in Ls(B)$. Then *the type* of l-formula z is such element $t \in Tp(S(B))$, denoted $tp(z)$, that $z\&t \in Ts(B)$.

Problems

1. What is the interpretation of the notion "a universe of formal study?"
2. What is the difference between l-formulas and t-formulas?
3. What is the difference between l-formulas and y-formulas?
4. What are K-strings?
5. What is the scheme of determining step by step three classes of formulas generated by a conceptual basis?
6. What is a K-calculus?
7. Describe the classes of formulas obtained due to employing only the rule $P[0]$.
8. What new manners of using the logical connectives $\wedge$ and $\vee$ (in comparison with the first-order predicate logic) are determined by the rule $P[7]$?
9. Why only one rule but not two rules govern the employment of the logical connectives $\wedge$ and $\vee$?

10. What rule enables us to join the referential quantifier to a simple or compound designation of a notion?
11. Describe the structure of compound designations of notions obtained as a result of employing the rule $P[8]$ at the last step of inference.
12. What rule is to be employed immediately after the rule $P[8]$ in order to obtain a compound designation of a concrete thing, event or a concrete set of things, events?
13. Infer (describe the steps of constructing) a K-string designating the set of all students of the Stanford University born in France. For this, introduce the assumptions about the considered conceptual basis.
14. Formulate the necessary assumptions about the considered conceptual basis and infer (describe the steps of constructing) the K-string:

$$(a)\ \ Subset(all\,car,\ all\,transport\_means)\,;$$

$$(b)\ \ (Manufacturer(certain\,car : x7) \equiv$$
$$certain\,company1 * (Name1, \text{“}Volvo\text{”}))\,;$$

$$(c)\ \ Belong((Barcelona \wedge Madrid \wedge Tarragona),$$
$$all\,city * (Country, Spain))\,;$$

$$(d)\ Belong(certain\,scholar * (First\_name, \text{“}James''\,)(Surname, \text{“}Hendler\text{”}) : x1,$$
$$Authors(certain\,inf\_object * (Kind1, sci\_article)$$
$$(Name1, \text{“}The\,Semantic\,Web\text{”})(Date\_of\_publ, 2001)))\,.$$