# Chapter 1
# Mathematical Models for Designing Natural Language Processing Systems as a New Field of Studies for Systems Science

**Abstract** This chapter grounds the necessity of developing new mathematical tools for the design of semantics-oriented natural language processing systems (NLPSs) and prepares the reader to grasping the principal ideas of these new tools introduced in the next chapters. Section 1.1 grounds the expedience of placing into the focus of Systems Science the studies aimed at constructing formal models being useful for the design of semantics-oriented NLPSs. Sections 1.2, 1.3, and 1.4 contain the proposals concerning the structure of such mathematical models of several new types that these models promise to become a great help to the designers of semantics-oriented NLPSs. Sections 1.5 and 1.6 jointly give the rationale for the proposed structure of new mathematical models. Section 1.5 states the central ideas of Cognitive Linguistics concerning natural language (NL) comprehension. Section 1.6 describes the early stage of the studies on formal semantics of NL. Section 1.7 sets forth the idea of developing formal systems of semantic representations with the expressive power close to that of NL; this idea is one of the central ones for this monograph.

## 1.1 An Idea of a Bridge Between Systems Science and Engineering of Semantics-Oriented Natural Language Processing Systems

Since the pioneer works of Montague [158–160] published in the beginning of the 1970s, the studies on developing formal semantics of natural language (NL) have been strongly influenced by the look at structuring the world suggested by mathematical logic, first of all, by first-order predicate logic.

However, it appears that a rich experience of constructing semantics-oriented natural language processing systems (NLPSs) accumulated since the middle of the 1970s provides weighty arguments in favor of changing the paradigm of formalizing

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_1,

NL semantics and, with this aim, placing this problem into the focus of interests of Systems Science.

The analysis allows for indicating at least the following arguments in favor of this idea:

1. The first-order logic studies the structure of propositions (in other words, the structure of statements, assertions). However, NL includes also the imperative phrases (commands, etc.) and questions of many kinds.
2. The engineering of semantics-oriented NLPSs needs, first of all, the models of transformers of several kinds. For instance, it needs the models of the subsystems of NLPSs constructing a semantic representation (SR) *Semrepr* of an analyzed sentence of an NL-discourse as the value of a function of the following arguments:
   - *Semcurrent* – a surface SR of the currently analyzed sentence *Sent* from the discourse *D*;
   - *Semold* – an SR of the left segment of *D* not including the sentence *Sent*;
   - *Lingbs* – a linguistic database, i.e., a collection of the data about the connections of lexical units with the conceptual (informational) units used by an algorithm of semantic-syntactic analysis of NL-texts; and
   - *Kb* – a knowledge base, or ontology, containing information about the world. But mathematical logic doesn't consider models of this kind.
3. NL-texts are formed as a result of the interaction of numerous mechanisms of conveying information acting in natural language. Mathematical logic doesn't possess formal means being sufficient for reflecting these mechanisms on a conceptual level. That is why mathematical logic doesn't provide sufficiently rich formal tools allowing for representing the results of semantic-syntactic analysis of arbitrary NL-texts. For instance, the first-order logic doesn't allow for building formal semantic analogues of the phrases constructed from the infinitives with dependent words, of sentences with the word "a notion," and of discourses with the references to the meanings of previous phrases and larger parts of discourses.

Due to these reasons, a hypothetical structure of several formal models of the new types for Systems Science is proposed in the next sections, expanding the content of [95]. In the examples illustrating the principal ideas of these new formal models, the strings belonging to formal languages of a new class will be used. It is the class of standard knowledge (SK) languages determined in Chapters 3 and 4 of this monograph. The strings of SK-languages will be employed for building semantic representations of word combinations, sentences, and discourses and for constructing knowledge modules. It is important to emphasize that it is not assumed that the reader of this chapter is acquainted with the definition of SK-languages. The purpose of the next sections is only to show the reasonability of undertaking the efforts for constructing the models of the proposed new types.

## 1.2 The Models of Types 1–4

This section proposes the structure of mathematical models of four new types promising to become useful tools of designing semantics-oriented NLPSs. In particular, this applies to the design of NL-interfaces to intelligent databases, autonomous intelligent systems (robots), advanced Web-based search engines, NL-interfaces of recommender systems, and to the design of NLPSs being subsystems of full-text databases and of the knowledge extraction systems.

### *1.2.1 The Models of Type 1*

The models of *the first proposed class* describe a correspondence between an introduced separate sentence in NL and its semantic representation. The transformation of the inputted sentences into their semantic representations is to be carried out with respect to a linguistic database *Lingb* and a knowledge base *Kb*.

Formally, the models of the proposed type 1 (see Fig. 1.1) describe a class of the systems of the form

$$(Linp, Lingbset, Kbset, Lsemrepr, transf, Alg, Proof),$$

where

- *Linp* is an input language consisting of sentences in natural language (NL);
- *Lingbset* is a set of possible linguistic databases (each of them is a finite set of some interrelated formulas);
- *Kbset* is a set of possible knowledge bases (each of them is also a finite set of some interrelated formulas);
- *Lsemrepr* is a language of semantic representations (in other terms, a language of text meaning representations);
- $transf$ is a mapping from the Cartesian product of the sets $Linp$, $Lingbset$, $Kbset$ to $Lsemrepr$;
- $Alg$ is an algorithm implementing the mapping (or transformation) $transf$;
- $Proof$ is a mathematical text being a proof of the correctness of the algorithm $Alg$ with respect to the mapping $transf$.

**Example 1.** If $Qs1$ is the question "How many universities in England use the e-learning platform "Blackboard" for distance education?" then $transf(Qs1)$ can be the string of the form

$$Question(x1, ((x1 \equiv Number1(S1)) \wedge Qualit-composition(S1,$$

$$university * (Location, England)) \wedge Description1(arbitrary\, university *$$

$$(Element, S1) : y1, Situation(e1, use1 * (Time, \#now\#)$$

$$(Agent1, y1)(Process, learning * (Kind1, online))$$
$$(Object1, certain\, platform3 * (Title, 'Blackboard'))))))).$$

This string includes, in particular, the following fragments: (a) a compound designation of the notion "an university in England," (b) a designation of arbitrary element of a set $S1$ consisting of some universities $arbitrary\, university * (Element, S1) : y1$, (c) a compound designation of an e-learning platform.

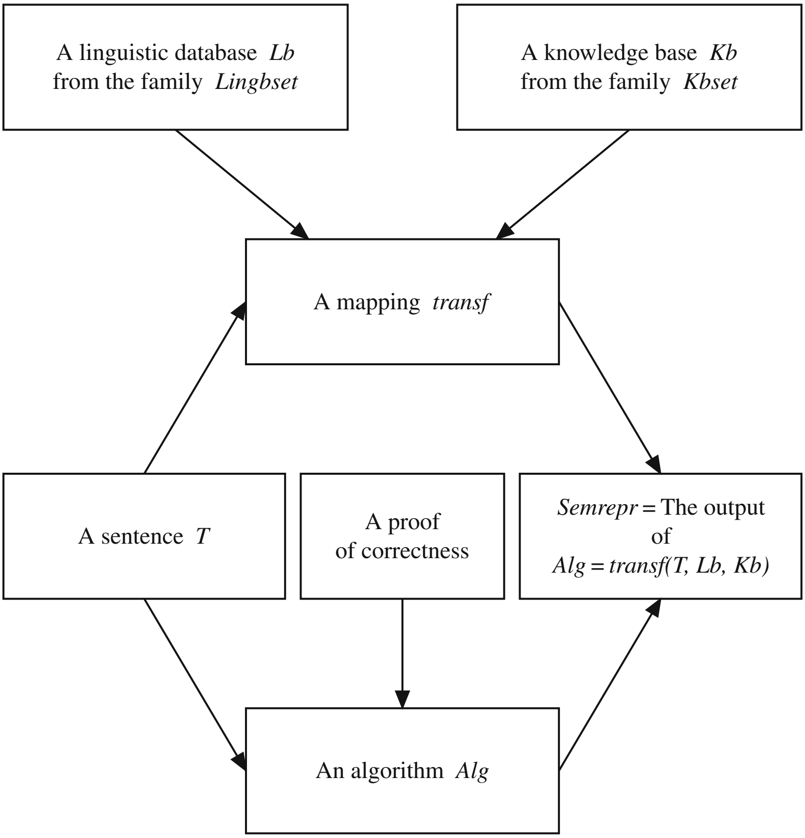


**Fig. 1.1** The structure of the models of type 1

The necessity of taking into account the component *Kb* (a knowledge base, an ontology) while formalizing the correspondence between NL-texts and their semantic representations follows from the analysis of numerous publications on semantics-oriented NL processing: from the pioneer works of Winograd [207], Wilks [205], Schank et al. [178, 179], Hobbs et al. [126] to, in particular, the papers [13, 17, 33, 154, 165, 208].

### 1.2.2 The Models of Type 2

The models of the *second proposed class* describe the systems of the form

$$(Linp, Lpatterns, Lingbset, Kbset, Loutput, transf, Alg, Proof),$$

where

- $Linp$ is an input language consisting of expressions in natural language (NL);
- $Lpatterns$ is a language intended for indicating the patterns for extracting information from inputted NL-texts;
- $Lingbset$ is the set of possible linguistic databases (each of them being a finite set of some interrelated formulas);
- $Kbset$ is the set of possible knowledge bases (each of them being also a finite set of some interrelated formulas);
- $Loutput$ is an output language;
- $transf$ is a mapping from the Cartesian product of the sets $Linp$, $Lpattern$, $Lingbset$, $Kbset$ to $Loutput$;
- $Alg$ is an algorithm implementing the mapping (or transformation) $transf$;
- $Proof$ is a mathematical text being a proof of the correctness of the algorithm $Alg$ with respect to the mapping $transf$.

**Example 2.** The model describes the work of a computer-intelligent agent looking in various business texts for information about any change in the world prices of aluminum or plumbum or copper for 3 or more percent. Here the language $Linp$ consists of NL-texts with commercial information, $Lpatterns$ contains a semantic representation (SR) of the expression "any change of the world prices for aluminium or plumbum or copper for 3 or more percents," and $Loutput$ consists of SRs of the fragments from inputted NL-texts informing about the changes of the world prices for aluminium or plumbum or copper for 3 or more percents. The string from the language $Lpatterns$ can have, for instance, the form

$$change1 * (Goods1, (aluminium \vee plumbum \vee copper))$$

$$(Bottom\_border, 3/percent)\,.$$

### 1.2.3 The Models of Type 3

Nowadays there are many known computer programs that are able to build semantic representations of separate short sentences in NL. However, there are a number of unsolved questions concerning the semantic-syntactic analysis of the fragments of discourses in the context of the preceding part of a dialogue or preceding part of a discourse. That is why it seems that the engineering of semantics-oriented NLPSs especially needs the models of the next proposed type.

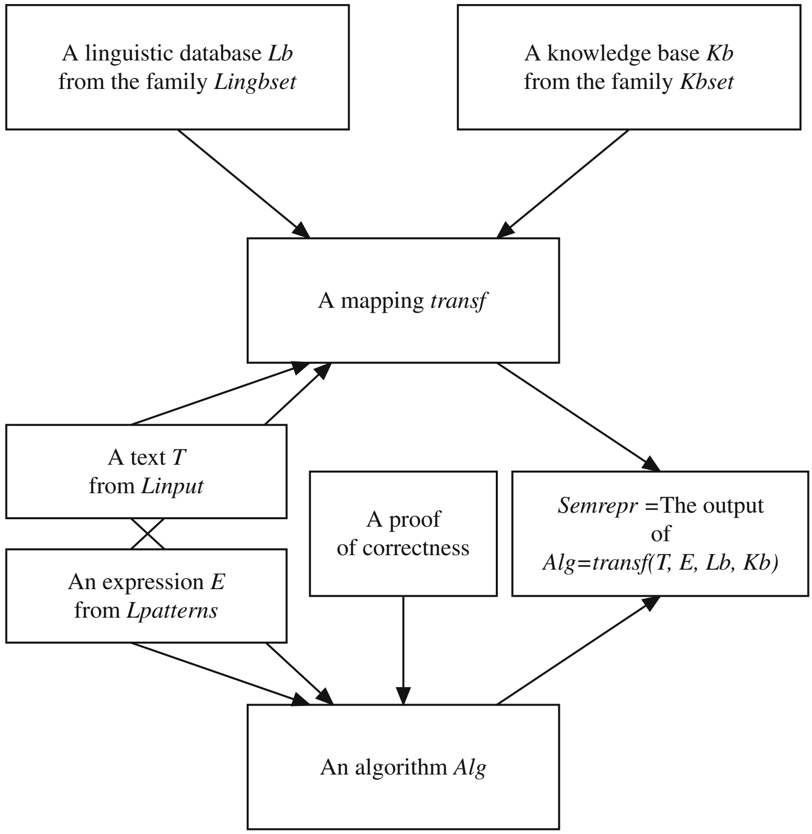


**Fig. 1.2** The structure of the models of type 2

The models of the *third proposed class* are intended for designing the subsystems of NLPSs interpreting a semantic representation of the current fragment of a discourse in the context of semantic representation of the preceding part of a dialogue or preceding part of a discourse.

Formally, the models of this class describe the systems of the form

$$(Lcontext, Linp, Lingbset, Kbset, Lsem, Lreact, transf, Alg, Proof),$$

where

- *Lcontext* is a language for representing the content of the preceding part of a dialogue or a discourse;
- *Linp* is an input language consisting of underspecified or completely specified semantic representations of NL-expressions (sentences and some fragments of sentences), such NL-expressions can be, in particular, the answers to the clarifying questions of a computer system;

- *Lingbset* and *Kbset* are (as above) the sets of possible linguistic databases and knowledge bases;
- the semantic language *Lsem* is intended for representing the deep meaning of the inputs from *Linp* with respect to a semantic representation of the preceding part of a dialogue or a discourse;
- *Lreact* is a language for building semantic descriptions of the computer system's reactions to the inputted texts;
- $transf$ is a mapping from the Cartesian product of the sets $Lcontext$, $Linp$, $Lingbset$, $Kbset$ to the Cartesian product of the sets $Lsem$ and $Lreact$;
- $Alg$ is an algorithm implementing the mapping (or transformation) $transf$;
- $Proof$ is a mathematical text being a proof of the correctness of the algorithm $Alg$ with respect to the mapping $transf$.

**Subclass 1:** The models describing the work of a Recommender System.

Since the beginning of the 2000s, a new branch of E-commerce has been quickly developing called Recommender Systems (RS). The software applications of this class are intended for consulting the end users of the Internet in order to help them to take decisions about the choice of goods and/or services. The key role in the functioning of many RS is the interaction with the users by means of Natural Language (NL) – English, German, etc. [23, 24, 111, 112].

Consider a particular interpretation of the components of the models of type 3 intended for the design of NL-interfaces to RS. An input $X1$ from $Lcontext$ reflects the content (in other words, the meaning) of the preceding part of a dialogue; an input $X2$ from *Linp* is an underspecified (or completely specified in particular cases) semantic representation (SR) of un utterance of the end user; an output $Y1$ from *Lsem* is a deep semantic representation of the input $X2$ in the context $X1$ with respect to a linguistic database *Lingbs* from the set *Lingbset* and to a knowledge base *Kbs* from the set *Kbset*; an output $Y2$ from *Lreact* is a semantic description of the computer system reaction to the inputted text with underspecified semantic representation $X2$ and deep semantic representation $Y1$. The knowledge base *Kbs* includes a subset of formulas $Userkbs$ interpreted as a User Model.

**Example 3.** Suppose that an end user of an RS of an Internet shop applies to the RS with the question $Qs1$ = "What models of the cell telephones of the firm Nokia do you have, the price from 300 USD to 450 USD?" Imagine that this question is transfomed into the semantic representation $X1$ of the form

$$Question(S1, Qualitative-composition(S1, model1*(Tech-product,$$

$$cell-telephone*(Manufacturer,\ \ certain\,firm1*(Name1, 'Nokia')$$

$$(Price-diapason, 300/USD, 450/USD)))).$$

Having received an answer to this question, the user can submit the next question Qs2 = "And of the firm Siemens?" It is an elliptical question, and the NL-interface to the discussed RS can transform $Qs2$ into the SR $X2$ of the form

$$Question(S2, Qualitative-composition(S2, techical-object*$$

$$(Manufacturer,\ \ certain\,firm1*(Name1,'Siemens'))).$$

In the English language, the question $Qs2$ can have one of the following two meanings in the context of the question $Qs1$:

- Meaning 1: The end user wants to get information about all available models of cell phones produced by the firm "Siemens";
- Meaning 2: The end user wants to get information about all available models of cell phones produced by the firm "Siemens" with the price from 300 to 450 USD.

That is why the model is to be constructed in such a way that the RS asks the end user to select one of these meanings. An SR of this question, denoted by $Clarif-qs$, belongs to the language $Lreact$.

Imagine that the end user selects the second meaning. Then, according to the model, the NL-interface to the RS forms the semantic representation *Semrepr* of the form

$$Question(S1,\ Qualitative-composition(S1,\ model1*(Tech-product,$$

$$cell-telephone*(Manufacturer,\ \ certain\,firm1*(Name1,'Siemens'))$$

$$(Price-diapason,\ 300/USD,\ 450/USD)))).$$

This string expresses the deep meaning of the question $Qs2$ in the context of the questions $Qs1$ with SR $X1$ and belongs to the language $Lsem$.

Of course, this example represents one of the simplest possible dialogues of a Recommender System with the end user. With respect to the achieved level of studies on NLPSs, many people today are able to elaborate a computer system being able to function in the described way. However, the real dialogues may be much more complex. That is why the practice of designing NLPSs really needs the models of the kind.

### *1.2.4 The Models of Type 4*

The models of *the fourth proposed class* are intended for designing computer-intelligent systems extracting knowledge from natural language sentences and complicated discourses for forming and updating a knowledge base (ontology) of an applied intelligent system. Such models describe the systems of the form

$$(Lcontext,\ Linp,\ Lingbset,\ Kbset,\ Lsem1,\ Lsem2,\ transf,\ Alg,\ Proof),$$

where

- $Lcontext$ is a language for building a semantic representation of the already processed part of an NL-text,

- *Linp* is an input language consisting of underspecified or completely specified semantic representations of NL-expressions (sentences and some fragments of sentences);
- *Lingbset* and *Kbset* are (as above) the sets of possible linguistic databases and knowledge bases;
- the semantic language $Lsem1$ is intended for representing the deep meaning of the inputs from *Linp* with respect to a semantic representation of the preceeding part of an NL-text;
- $Lsem2$ is a language for representing the knowledge of the required kinds extracted from the expressions of the language $Lsem1$;
- $transf$ is a mapping from the Cartesian product of the sets $Lcontext$, $Linp$, $Lingbset$, $Kbset$ to the Cartesian product of the sets $Lsem1$ and $Lsem2$;
- $Alg$ is an algorithm implementing the mapping (or transformation) $transf$; and
- $Proof$ is a mathematical text being a proof of the correctness of the algorithm $Alg$ with respect to the mapping $transf$.

## 1.3 The Models of Type 5

The models of *the fifth proposed class* are the models of advanced question answering systems, i.e., the models of intelligent computer systems being able to find an answer to a request in NL of an end user of a full-text database (of course, it can be Web-based) as a result of semantic-syntactic analysis of NL-texts stored in this database. Such models take into account the fact that the information enabling an intelligent system to formulate an answer to the posed question can be accumulated step by step, in the course of processing several texts in different informational sources.

The models of *the fifth proposed class* describe the systems of the form

$$(Lreq, Lct, Linp, Lbset, Kbset, Ls1, Ls2, Lans, Ind, transf, Alg, Proof),$$

where

- *Lreq* is a language for constructing semantic representations of the requests posed by the end users of an intelligent system;
- *Lct* is a language for building a semantic representation of the already processed part of an NL-text;
- *Linp* is an input language consisting of underspecified or completely specified semantic representations of NL-expressions (sentences and some fragments of sentences and discourses);
- *Lbset* and *Kbset* are (as above) the sets of possible linguistic databases and knowledge bases;
- the semantic language $Ls1$ is intended for representing the deep meaning of the inputs from *Linp* with respect to a semantic representation of the preceeding part of an NL-text;

- $Ls2$ is a language for representing a piece of knowledge to be inscribed into the knowledge base $Kb$ in order to be used later for formulating an answer to the request of the end user;
- $Lans$ is a language for expressing the answers to the input requests, it includes the symbol *nil* called "the empty answer";
- $Ind = \{0, 1\}$, where 0 is interpreted as the signal to stop the search, 1 is interpreted as the signal to continue the search;
- $transf$ is a mapping from the Cartesian product of the sets $Lreq$, $Lct$, $Linp$, $Lbset$, $Kbset$ to the Cartesian product of the sets $Ls1$, $Ls2$, $Lans$, $Ind$;
- $Alg$ is an algorithm implementing the mapping (or transformation) $transf$;
- $Proof$ is a mathematical text being a proof of the correctness of the algorithm $Alg$ with respect to the mapping $transf$.

Continuing the line of [82], let's illustrate some desirable properties of formal models reflecting the basic mechanisms of the hypothetical computer-intelligent systems of the kind.

Imagine that there is a big city D., and a user of an intelligent full-text database $Db1$ inputs the question $Qs$ = "Is it true that the ecological situation in the city D. has improved during the year?" and the date of inputting $Qs$ is $Date1$.

Suppose that $Qs$ is transformed into the following initial semantic representation $Semreprqs1$ :

$$Question(u1, (u1 \equiv Truth - value(Better(Ecology(certain\,city*$$

$$(Name1, 'D.') : x1, Year(Date1)), Ecology(x1, Last - year(Date1)))))).$$

In the expression $Semreprqs1$, the element $Ecology$ is to be interpreted as the name of the function assigning to the space object $z1$ and the year $z2$ a statement about the ecological situation in $z1$ corresponding to $z2$.

Let's assume that $Db1$ has the knowledge base $Kb1$ including a part $Objects - list$, and this part contains the string $certain\,city * (Name1, \text{“}D\text{”}.) : v315$. This means that the city D. is associated with the variable $v315$, playing the role of the unique system name of this city. Suppose also that $Date1$ corresponds to the year 2008. Then $Semreprqs1$ is transformed into the secondary semantic representation $Semreprqs2$ of the form

$$Question(u1, (u1 \equiv Truth - value(Better(Ecology(certain\,city*$$

$$(Name1, 'D.') : v315, 2008), Ecology(v315, 2007))))).$$

Suppose that there is the newspaper "D. News," and one of its issues published in the same month as $Date1$ contains the following fragment $Fr1$ : *"The quantity of species of birds who made their nests in the city has reached the number 7 in comparison with the number 5 a year ago. It was announced at a press-conference by Monsieur Paul Loran, Chair of the D. Association of Colleges Presidents"*.

Let's consider a possible way of extracting from this fragment the information for formulating an answer to $Qs$. The first sentence $Sent1$ of $Fr1$ may have the following SR $Semrepr1a$ :

$$((Quantity(certn\,species*(Compos1,\,bird)(Descr1,\,\langle S1,\,P1\rangle))\equiv 7)\wedge$$
$$(P1\ \equiv\ (Compos1(S1,\,bird)\wedge Descr2(arbitrary\,bird*$$
$$(Elem,\,S1)\ :\ y1,\,\exists e1\,(sit)\,Is1(e1,\,nesting*$$
$$(Agent1,\,y1)(Loc,\,x1)(Time,\,2008)))))\wedge$$
$$((Quantity(certn\,species*(Compos1,\,bird)(Descr1,\,\langle S2,\,P2\rangle))\equiv 5)\wedge$$
$$(P2\ \equiv\ (Compos1(S2,\,bird)\wedge Descr2(arbitrary\,bird*$$
$$(Elem,\,S2)\ :\ y2,\,\exists e2\,(sit)\,Is1(e2,\,nesting*$$
$$(Agent1,\,y2)(Loc,\,x1)(Time,\,2007))))))\ :\ P3\,.$$

The symbol $certn$ is the informational unit corresponding to the word combination "a certain"; $Compos1$ is the designation of the binary relation "Qualitative composition of a set"; $P1$, $P2$, $P3$ are such variables that their type is the distinguished sort "a meaning of proposition".

Suppose that the second sentence $Sent2$ of $Fr1$ has the following semantic representation $Semrepr2a$ :

$$\exists e3\,(sit)\,(Is(e3,\,announcing*(Agent1,\,x2)(Content1,\,P3)(Event,$$
$$certn\,press-conf\ :\ x3))\wedge(x2\equiv certain\,man*(First-name,\,\text{“}Paul''\text{)}$$
$$(Surname,\,'Loran'))\wedge(x2\equiv Chair(certn\,association1*$$
$$(Compos1,\,scholar*(Be1,\,President(any\,college*(Location,$$
$$certn\,city*(Name1,\,\text{“}D\text{”.})\ :\ x4))))))).$$

Here the element $association1$ denotes the concept "association consisting of people" (there are also the associations of universities, cities, etc.).

The analysis of the first sentence $Sent1$ shows that it is impossible to find directly in $Sent1$ the information determining the referent of the word "the city." In this case, let's take into account the knowledge about the source containing the fragment $Fr1$ and about the use of this knowledge for clarifying the referential structure of published discourses.

Imagine that the knowledge base $Kb1$ of the considered hypothetical intelligent system contains the string

$$Follows(((z1\equiv arbitrary\,edition*(Title,\,z2)(Content1,\,Cont1))\wedge$$
$$Associated(z2,\,arbitrary\,space-object\ :\ z3)\wedge Element(w,\,pos,\,Cont1)\wedge$$
$$Sem-class(w,\,pos,\,space-object)\wedge No-show-referent(w,\,pos,\,Cont1)),$$
$$Referent(w,\,pos,\,Cont1,\,z3)).$$

Let's interpret this formula as follows. Suppose that: (1) an arbitrary edition $z1$ has the title $z2$ and the content $Cont1$, its title $z2$ is associated in any way with the space-object $z3$; (2) the string $Cont1$ contains the element $w$ in the position $pos$, its semantic class is $space - object$ (a city, a province, a country, etc.); (3) the text contains no explicit information about the referent of the element $w$ in the position $pos$ of the formula $Cont1$. Then the referent of the element $w$ is the entity denoted by $z3$.

In order to use this knowledge item for the analysis of the fragment $Fr1$, let's remember that the list of the objects $Objects - list$ (being a part of the knowledge base $Kb1$) includes the string

$$certn\,city * (Name1, 'D.') \,:\, v315.$$

Then the system transforms the semantic representation $Semrepr1a$ of the first sentence $Sent1$ into the formula $Semrepr1b$ of the form

$$(Semr1a \wedge (x1 \equiv v315)).$$

This means that at this stage of the analysis the information extracted from $Sent1$ is associated with the city D.

Assume that the knowledge base $Kb1$ contains the knowledge items

$$\forall z1\,(person)\,\forall c1\,(concept)\,Follows(Head(z1,$$

$$arbitrary\,association1 * (Compos1, c1)), Is1(z1, c1)),$$

$$\forall z1\,(person)\,Follows((z1 \equiv$$

$$President((arbitrary\,univ \,:\, z2 \vee arbitrary\,college \,:\, z3))),$$

$$Qualification(z1, Ph.D.)),$$

and these items are interpreted as follows: (1) if a person $z1$ is the head of an association of the type 1 (associations consisting of people), the concept $c1$ qualifies each element of this association, then $z1$ is qualified by $c1$ too; (2) if a person $z1$ is the president of a university or a college, $z1$ has at least a Ph.D. degree.

Proceeding from the indicated knowledge items and from $Semrepr2a$, the system builds the semantic representation $Semrepr2b$ of the form

$$(Semr2a \wedge (x1 \equiv v315))$$

and then infers the formula $Qualification(x2, Ph.D.)$, where the variable $x2$ denotes Monsieur Paul Loran, Chair of the D. Association of Colleges Presidents.

Let $Kb1$ contain also the expression

$$Follows\,(\exists e1\,(sit)\,Is1(e1, announcing * (Agent1, arbitrary\,scholar*$$

$$(Qualif, Ph.D.))(Kind - of - event, personal - communication)$$

$$(Content1, Q1)(Time, t1)), Truth - estimation(Q1, t1, \langle 0.9, 1 \rangle)),$$

interpreted as follows: if a scholar having a Ph.D. degree announces something, and it is not a personal communication then the estimation of the truth of the announced information has a value in the interval [0.9, 1.0]. Here the substring $\exists e1\,(sit)\,Is(e1, announcing*$ is to be read as "There is an event $e1$ of the type 'announcing' such that."

So let's imagine that, proceeding from the semantic representations $Semrepr1b$ and $Semrepr2b$ (the secondary semantic representations of the first and second sentences of the fragment $Fr1$) and the mentioned knowledge items from $Kb1$, the system infers the expression

$$((Quantity(certn\,species * (Compos1, bird)(Descr1, \langle S1, P1 \rangle)) \equiv 7) \wedge$$
$$(P1 \equiv (Compos1(S1, bird) \wedge Descr2(arbitrary\,bird*$$
$$(Elem, S1) : y1, \exists e1\,(sit)\,Is1(e1, nesting*$$
$$(Agent1, y1)(Loc, v315)(Time, 2008))))) \wedge$$
$$((Quantity(certn\,species * (Compos1, bird)(Descr1, \langle S2, P2 \rangle)) \equiv 5) \wedge$$
$$(P2 \equiv (Compos1(S2, bird) \wedge Descr2(arbitrary\,bird*$$
$$(Elem, S2) : y2, \exists e2\,(sit)\,Is1(e2, nesting*$$
$$(Agent1, y2)(Loc, v315)(Time, 2007)))))) : P3\,.$$

Suppose that the knowledge base $Kb1$ contains the knowledge unit

$$\forall z1(space - object)\,\forall t1(year)\,Follows(Better(Ecolog - sit(z1, bird, t1),$$

$$Ecolog - sit(z1, bird, t2)), Better(Ecology(z1, t1), Ecology(z1, t2)))$$

and the knowledge unit

$$\forall z1(space.ob)\forall t1(year)\forall t2(year)\,Follows((Compos1(S1, bird) \wedge$$

$$Descr2(arbitrary\,bird * (Elem, S1) : y1,$$
$$\exists e1\,(sit)\,Is1(e1, nesting*$$
$$(Agent1, y1)(Loc, z1)(Time, t1))) \wedge (Compos1(S2, bird)$$
$$\wedge Descr2(arbitrary\,bird * (Elem, S2) : y2,$$
$$\exists e2\,(sit)\,Is1(e2, nesting*$$
$$(Agent1, y2)(Loc, z1)(Time, t2))) \wedge$$
$$Greater\,(Quantity(S1), Quantity(S2))),$$
$$Better(Ecolog - sit(z1, bird, t1),$$

$$Ecolog - sit(z1, bird, t2))).$$

That is why the system finally infers the formulas

$$Better\,(Ecolog - sit(v315, bird, 2008), Ecolog - sit(v315, bird, 2007)),$$

$$Better\,(Ecology(v315, 2008), Ecology(v315, 2007)).$$

Hence the system outputs the expression of the form

$$\langle \text{“}YES\text{”}, Ground: Fr1\,('D.News', Date1)\rangle\,.$$

## 1.4 The Significance of the Models for the Design of Linguistic Processors

The analysis shows the significance of the studies aimed at constructing the formal models of the considered kinds for the engineering of natural language processing systems (NLPSs). In particular, the following factors are distinguished:

1. The algorithms being components of such formal models can be directly used as the algorithms of the principal modules of NLPSs.
2. The descriptions of the mappings $transf$, characterizing the correspondence between the inputs and outputs of the systems, can become the principal parts of the documentation of such programming modules. As a result, the quality of the documentation will considerably increase.
3. The designers of NLPSs will get the comprehensible formal means for describing the semantics of lexical units and for building semantic representations of complicated natural language sentences and discourses in arbitrary application domains. This will contribute very much to the transportability of the elaborated software of NLPSs as concerns new tasks and new application domains.

It should be underlined that even the elaboration of the *partial models* of the kind promises to be of high significance for the engineering of NLPSs. The principal difference between the complete models and partial models of the considered types consists in the lack of a proof of the correctness of the algorithm $Alg$ with respect to the defined transformation $transf$. Besides, a partial model may include a not mathematically complete definition of a transformation $transf$ but only a description of some principal features of this transformation.

Even in case of partial models, the designers of semantics-oriented NLPSs will receive an excellent basis for the preparation of such documentation of a computer system that distinguishes the most precious or original features of the algorithms and/or data structures and creates the good preconditions of transporting the data structures and algorithms to new problems and application domains.

## 1.5 The Context of Cognitive Linguistics for the Formal Study of Natural Language

The first bunch of ideas underlying the proposed structure of the mathematical models intended for the design of semantics-oriented NLPSs was given by the analysis of the look of Cognitive Science at the regularities of natural language comprehension.

The problems and achievements in the field of constructing NLPSs, on the one hand, and huge difficulties in the way of formalizing regularities of NL-comprehension, on the other hand, evoked in the 1980s a considerable interest of many psychologists and linguists in investigating such regularities. A new branch of linguistics was formed called Cognitive Linguistics as a part of Cognitive Science.

Cognitive linguists consider language as "an instrument for organizing, processing, and conveying information. The formal structures of languages are studied not as if they were autonomous, but as reflections of general conceptual organization, categorization principles, processing mechanisms, and experiential and environmental influences" (see [103], p. 1).

The obtained results allowed for formulating in [66] the following now broadly accepted principles of natural language comprehension.

1. The meaning of a natural-language text (NL-text) is represented by means of a special mental language, or a language of thought [11, 20, 22, 39, 49, 50, 135, 139, 149–152, 155, 178, 203].
2. People build two different (though interrelated) mental representations of an NL-text. The first one is called by Johnson-Laird [139] the *prepositional representation (PR).* This representation reflects the semantic microstructure of a text and is close to the text's surface structure.
   The second representation being a mental model (MM) is facultative. The MM of a text reflects the situation described in the text. Mental models of texts are built on the basis of both texts' PRs and diverse knowledge – about the reality, language, discussed situation, and communication participants [139].
3. A highly important role in building the PRs and MMs of NL-texts is played by diverse cognitive models accumulated by people during the life-semantic frames, explanations of notions' meanings, prototypical scenarios, social stereotypes, representations of general regularities and area-specific regularities, and other models determining, in particular, the use of metaphors and metonymy [39, 45, 137, 139, 149, 150, 156, 187].
4. The opinion that there exists syntax as an autonomous subsystem of language system has become out of date. Syntax should depend on descriptions of cognitive structures, on semantics of NL. Natural language understanding by people doesn't include the phase of constructing the pure syntactic representations of texts.
   The transition from an NL-text to its mental representation is carried out on the basis of various knowledge and is of integral character [20, 48, 139, 149, 151, 152, 187, 198].

5. Semantics and pragmatics of NL are inseparably linked and should be studied and described by the same means [180, 189].

A significant role in formulating the enumerated principles was played by the researches on developing computer programs being capable of carrying out the conceptual processing of NL-texts. This applies especially to the studies that can be attributed to the semantics-oriented (or semantically driven) approaches to natural language parsing.

It appears that the set of principles stated above may serve as an important reference-point for the development and comparison of the available approaches to mathematical modeling of NL-understanding.

## 1.6 Early Stage of Natural Language Formal Semantics

The shortcomings of the main-known approaches to the formal study of NL-semantics were felt in the 1980s by many philosophers, psychologists, and linguists. The basic philosophical ideas of model-theoretic semantics were criticized, in particular, by Putnam [170], Johnson-Laird [139], Fillmore [45], Seuren [187, 188], and Lakoff [149].

The main approaches to the formalization of NL-semantics popular in the 1980s – Montague Grammar and its extensions [141, 158–160, 164, 166], Situation Semantics [2, 15, 16, 42], Discourse Representation Theory [122, 142–144], and Theory of Generalized Quantifiers [14] – are strongly connected with traditions of mathematical logic, of model-theoretic semantics, and do not provide formal means permitting to model the processes of NL-comprehension in correspondence with the above enumerated principles of cognitive science.

In particular, these approaches do not afford effective formal tools to build (a) semantic representations of arbitrary discourses (e.g., of discourses with the references to the meaning of fragments being sentences or larger parts of texts), (b) diverse cognitive models, for instance, explanations of notions' meanings, representations of semantic frames, (c) descriptions of sets, relations and operations on sets.

Besides, these approaches are oriented toward regarding assertions. However, it is also important to study the goals, commitments, advices, commands, questions.

The dominant paradigm of describing surface structure of sentences separately from describing semantic structure (stemming from the pioneer works of Montague [158–160]) contradicts one of the key principles of cognitive linguistics – the principle assuming the dependency of syntax in semantics.

Highly emotionally the feeling of dissatisfaction with the possibilities of the main popular approaches to the formalization of NL-semantics was put into words by Seuren [187, 188]. In particular, Seuren expressed the opinion that the majority of studies on the formalization of NL-semantics was carried out by researchers interested, first of all, in demonstrating the use of formal tools possessed by them, but not in developing the formal means allowing for modeling the mechanisms of NL-comprehension.

As it is known, ecology studies the living beings in their natural environment. In [188], the need of new, adequate, ecological approaches to studying the regularities of NL-comprehension is advocated. Many considerations and observations useful for working out ecological approaches to the formalization of discourses' semantics were formulated by Seuren. In the monograph [188], a peculiar attention was given to the questions of expressing and discerning the presuppositions of discourses, and the so-called Presuppositional Propositional Calculus was introduced.

In the second half of the eighties, a number of new results concerning the formalization of NL-semantics was obtained. Let us mention here the approach of Saint-Dizier [177] motivated by the tasks of logic programming, the results of Cresswell [29] and Chierchia [26] on describing structured meanings of sentences, the theory of situation schemata [42], Dynamic Semantics in the forms of Dynamic Predicate Logic [108] and Dynamic Montague Grammar [34, 106, 107].

Unfortunately, the restrictions pointed above in this section apply also to these new approaches. It should be added that Chierchia [26] describes structured meanings of some sentences with infinitives. But the expressive power of semantic formulae corresponding to such sentences is very small in comparison with the complexity of real discourses from scientific papers, text books, encyclopedic dictionaries, legal sources, etc.

Thus, the approaches mentioned in this section do not provide effective and broadly applicable formal tools for modeling NL-understanding in accordance with stated principles of cognitive psychology and cognitive linguistics.

The lack in the eighties of such means for modeling NL-understanding can be seen also from the substantial text book on mathematical methods in linguistics by Partee et al. [167].

One can't say that all approaches to the formalization of NL-semantics mentioned in the precedent subsection are not connected with the practice of designing NLPSs. There are publications, for example, on using for the design of NLPSs Montague Grammar in modified forms [28, 125, 185], Situation Semantics [213], and Discourse Representation Theory [124].

The language of intentional logic provided by Montague Grammar is used also in Generalized Phrase Structure Grammars [102] for describing semantic interpretations of sentences. Such grammars have found a number of applications in natural language processing.

Nevertheless, these and other approaches mentioned in this section possess a number of important shortcomings as concerns applying formal methods to the design of NLPSs and to developing the theory of NLPSs.

The demands of diverse application domains to the formal means for describing natural language may differ. That is why we distinguish for further analysis the following groups of application domains:

- Natural-language interfaces to databases, knowledge bases, autonomous robots.
- Full-text databases; computer systems automatically forming and updating knowledge bases of artificial intelligence systems by means of extracting information from scientific papers, text books, etc., in particular, intelligent text summarization systems.

- Such subsystems of automatized programming systems that are determined for transforming the NL-specifications of tasks into the formal specifications for the further synthesis of computer programs; such similar subsystems of computer-aided design systems that are determined for transforming the NL-specifications of technical objects to be designed into the formal specifications of these objects.

Obviously, the enumerated application domains represent only a part of all possible domains, where the development and use of NLPSs are important. However, for our purpose it is sufficient to consider only the mentioned important domains of applying NLPSs.

The analysis of formal means for the study of NL needed for these domains will allow us to get a rather complete list of demands to the formal theories of NL which should be satisfied by useful for practice and broadly applicable mathematical tools of studying NL-semantics and NL-pragmatics.

Let us regard for each distinguished group of applications the most significant restrictions of the approaches to the formal study of NL-semantics mentioned in this section.

**Group 1**

Semantics-oriented, or semantically driven NL-interfaces work in the following way. They transform an NL-input (or at first its fragment) into a formal structure reflecting the meaning of this input (or the meaning of a certain input's fragment) and called a semantic representation (SR) or a text meaning representation of the input or input's fragment. Then the SR is used (possibly, after transforming into a problem-oriented representation) for working out a plan for the reaction to the input with respect to a knowledge base, and after this a certain reaction is produced. The reactions may be highly diverse: applied intelligent systems can pose questions, fulfill calculations, search for required information, and transport things.

For constructing NL-interfaces in accordance with these principles, the following shortcomings of MG and its extensions, including Dynamic Montague Grammar, of Situation Semantics, Discourse Representation Theory, Theory of Generalized Quantifiers, Dynamic Predicate Logic, and of other approaches mentioned in this section are important:

1. The effective formal means for describing knowledge fragments and the structure of knowledge bases are not provided; in particular, this applies to the formal means for building semantic representations of complex definitions of notions.
2. There are no sufficiently powerful and flexible formal means to describe surface and semantic structures of questions and commands expressed by complicated NL-utterances.
3. There are no sufficiently powerful and flexible formal means to represent surface and semantic structures of the goals of intelligent systems expressed by complex NL-utterances.
4. The possibilities of intelligent systems to understand the goals of communication participants and to use the information about these goals for planning the reaction to an NL-input are not modeled.

5. The enumerated approaches do not give the flexible and powerful formal means for describing structured meanings of NL-discourses (including the discourses from scientific papers, legal sources, patents, etc.). The means of describing structured meanings of discourses are extremely restricted and unsatisfactory from the viewpoint of practice. In particular, the discourses with the references to the meaning of sentences and larger fragments of texts are not considered.
6. The existence of sentences of many types broadly used in real life is ignored. For instance, the structure of the following sorts of sentences is not studied: (a) containing expressions built from the descriptions of objects, sets, notions, events, etc., by means of logical connectives ("Yves has bought a monograph on mathematics, a text-book on chemistry, and a French-Russian dictionary"), (b) describing the operations on sets ("It will be useful to include Professor A. into the Editorial Board of the journal B."), (c) with the words "notion" or "term" (the latter in the meaning "a notion").
7. The models of the correspondences between the texts, knowledge about the reality, and semantic representations of texts are not built, and adequate means for developing models of the kind are not provided.
8. The inputs of NLPSs may be incomplete phrases, even separate words (for instance, the answers to questions in the course of a dialogue). The interpretation of such inputs is to be found in the context of precedent phrases and with respect to the knowledge about the reality and about the concrete discussed situation. However, such a capability of NL-interfaces isn't studied and isn't formally modeled.
9. The structure of metaphors and incorrect but understandable expressions from input texts, the correspondences between metaphors and their meanings are not investigated by formal means.

Wilks ([206], p. 348) writes that many NLPSs (in particular, the systems of machine translation) do not work as explained by the "official" theories in publications about these systems and function "in such a way that it cannot be appropriately described by the upper-level theory at all, but requires some quite different form of description." The analysis carried out above shows that the approaches mentioned in this section do not afford the opportunities to adequately describe the main ways of processing information by semantic components of NLPSs.

**Group 2**

Obviously, the restrictions 1, 3–6, and 8–9 are also important from the viewpoint of solving tasks like the development of intelligent full-text databases. The restriction 8 should be replaced by a similar restriction, since the fragments of discourses pertaining to business, technology, science, etc., may be incomplete, elliptical phrases.

The following restriction is to be pointed out additionally: the semantic structure of discourses with the proposals, commitments (the protocols, contracts often include such discourses), the interrelations between surface and semantic structures are not studied and modeled.

Over thirty years ago Wilks ([205], p. 116) noted that "any adequate logic must contain a dictionary or its equivalent if it is to handle anything more than terms with naive denotations such as 'chair'."

However, all approaches to the formalization of NL-semantics mentioned in this section do not take into account the existence and the role of various conceptual dictionaries. Due to this, in particular, reason, there is no opportunity to model the correspondence between texts, knowledge about the reality, and semantic representations of texts.

At first sight, the demands to the means of describing structured meanings of discourses and to the models of the correspondences between texts, knowledge, and SRs of texts are much stronger for the second group of applications than for the first one.

Nevertheless, it is not excluded that the joint future work of philosophers, linguists, specialists on computer science, and mathematicians will show that such demands are in fact very similar or the same for these two groups of NLPSs' applications.

**Group 3**

In addition to the shortcomings important for the groups 1 and 2, the following restriction should be mentioned: there are no effective formal means to represent structured meanings of NL-discourses describing the algorithms, the methods of solving diverse tasks. In particular, there are no adequate formal means to describe on semantic level the operations with the sets.

It appears that the collection of restrictions stated above provides a useful reference-point for enriching the stock of the means and models for the mathematical study of NL-communication.

## 1.7 The Significance of Highly Expressive Formal Systems of Semantic Representations

The collection of the tasks faced by the theory of semantics-oriented linguistic processors (LPs) in the beginning of the 1980s proved to be extremely complicated. As a consequence, the development of the theory of LPs in the 1980s slowed down. Though many projects of designing LPs were fulfilled in a number of countries, a substantial progress in this field was not achieved.

The principal cause of this deceleration is as follows. In natural language, numerous mechanisms of coding and decoding information interact in an intricate manner. That is why in order "to understand" even rather simple (for the human being) phrases and discourses, a computer system very often has to use the knowledge about the regularities of different levels of NL (morphological, syntactical, semantic) as well as the knowledge about thematic domains and the concrete situation of communication.

For instance, for making the decision about the referent of the pronoun "them", it may be necessary to apply common sense and/or to carry out a logical reasoning. Similar situations take place for the problem of reconstructing the meaning of an elliptical phrase (i.e., a phrase with some omitted words and word combinations) in the context of a discourse or a communicative situation.

That is why, while trying to formalize the understanding by a computer system even of rather simple NL-texts, the researchers quickly came to the conclusion that for solving their particular tasks, it is necessary to first find the theoretical decisions pertaining to arbitrary texts from a group of natural languages (for example, English, German, French). As a consequence, in the 1980s, in the scientific publications on NL processing even the metaphor "the theory bottleneck" emerged as a reflection of the considerable obstacles to be overcome for creating an adequate theory of understanding NL by computer systems.

Fortunately, several groups of the researchers from different countries (including the author of this monograph) proposed an idea that allowed finding a way out of the described deadlock situation. The essence of this idea is as follows. It is necessary to elaborate such formal languages for representing knowledge about the world and for building semantic representations (SRs) of NL-texts that these languages provide the possibility to construct SRs as the formal expressions reflecting many structural peculiarities of the considered NL-texts.

In other words, it is necessary to develop the formal languages (or formal systems, since the set of well-formed expressions given by the definition of a formal system is a formal language) for describing structured meanings of NL-texts with the expressive possibilities being rather close to the expressive possibilities of NL.

In this case it will be possible to carry out the semantic-syntactic analysis of an NL-text T from the considered sublanguage of NL in two stages reflected by the scheme

$$A\ NL-text\ T \Rightarrow Underspecified\ semantic\ representation\ of\ T$$

$$\Rightarrow Final\ (completely\ specified)\ semantic\ representation\ of\ T.$$

This scheme is to be interpreted in the following way. First, an intermediary semantic representation of the analyzed text $T$ is to be constructed, it is called an underspecified semantic representation (USR) of $T$. Most often, this expression will reflect only partially the meaning of the considered text $T$. For instance, an USR of the input text $T$ may indicate no referent of the pronouns "her" or "them" from the text $T$ or may indicate no concrete meaning of the word "station" from $T$ but only the set of all possible meanings of this word.

However, an USR of an NL-text $T$ is an formal expression in contrast to the text $T$. That is why during the second stage of processing $T$ it will be possible for eliminating an uncertainty to call and apply one of numerous specialized procedures being "experts" on concrete questions. Such procedures can be developed with the use of formal means of representing information, because the databases and knowledge bases of linguistic processors store the expressions of formal knowledge representation languages, and the USR to be analyzed and the final SR of the input text $T$ are formal expressions.

It must be noticed that this idea was formulated for the first time in the author's works [52–56], and the new classes of formal languages of semantic representations with very high expressive power were defined in the mentioned works (see next sections).

Since the end of the 1980s, the idea of employing the formal systems of semantic representations with high expressive possibilities in the design of semantics-oriented LPs has been the central one for the development of the theory of understanding NL by computer-intelligent systems. The growth of the popularity of this idea was stimulated in the 1980s and 1990s by

- the series of publications on Episodic Logic (EL) [130–134, 181–184], the use of EL as the theoretical basis for the implementation of the project TRAINS, aimed at the formalization of problem-oriented dialogue in natural (English) language [4];
- the realization of the machine translation project Core Language Engine (CLE) in the Cambridge division (England) of the Stanford Research Institute [8, 9];
- the implementation in the 1980s–1990s of the project SnepS in USA [192, 193];
- the publications on the Theory of Conceptual Graphs [37, 195–197].

The general feature of the major part of the proposed approaches to constructing formal systems of semantic representations with high expressive power can be characterized as enhancing the expressive possibilities of first-order predicate logic by means of adding a number of new possibilities reflecting (on the formal level) some expressive mechanisms functioning in natural language.

**Example 1.** We often encounter in discourses the fragments "due to this event," "this caused," and the like. The referent of the word combination "this event" (first fragment) and of the pronoun "this" (second fragment) is a situation which took place at some moment in the past. NL allows for using in discourses such short designations of the situations in case a previous fragment of the considered discourse contains a complete description of this situation. It is a manifestation of one of the mechanisms of compactly coding information in NL.

However, the first-order predicate logic provides no possibility to associate an arbitrary formula $F$ being a part of a formula $H$ with a mark (being a variable or a constant) and then use only this compact mark instead of all other occurrences of $F$ in $H$.

Episodic Logic overcomes this restriction of first-order predicate logic, and it is one of EL's distinguished features. Suppose, for instance, that $T1$ = "A predatory animal attacks a nearby creature only when it is hungry or feeling nasty." Then $T1$ may have the following semantic representation [130]:

$$(\forall x : \ [x((attr\ predatory)\ animal)]((\forall y : \ [y\ creature]$$

$$(\forall e1 : \ [[y\ near\ x] **e1]$$

$$(\forall e2 : \ [e2\ during\ e1]$$

$$[[[x\ attack\ y] **e2]$$

$$(\exists e3 : \ [e3\ same-time\ e2]$$

$$[[[x\ hungry] **e3] \vee [[x\ feel-nasty] **e3]])])]))).$$

In this formula of EL, the string $**$ designates the episodic operator; this operator connects a formula with the mark of the situation (or episode) it describes. The introduction of the episodic operator provides the possibility to model the mechanism of compactly encoding information in NL manifested in discourses due to the word combinations "this event," "this situation," "this caused," etc.

**Example 2.** Sentences and discourses in NL often contain compound designations of sets. For instance, commercial contracts may contain the expressions "5 containers," "a party of containers with bicycles," etc. The first-order predicate logic gives no convenient means for building formal analogues of the compound designations of sets.

In the project Core Language Engine (CLE), the formal expressions used for constructing underspecified semantic representations of the sentences are called quasi-logical forms (QLF). In particular, the expression "the three firms" can be associated with the QLF (see [6])

$$q\_term(\langle t = quant,\ n = plur,\ l = all\rangle),$$

$$S,$$

$$[subset,\ S,$$

$$q\_term(\langle t = ref,\ p = def,\ l = the,\ n = number(3)\rangle,$$

$$X,\ [firm,\ X])]).$$

Thus, the language of quasilogical forms allows for building formal analogues of the compound natural language designations of sets.

Analysis shows that none of the approaches to the formalization of semantic structure of NL-texts mentioned above in this section is convenient for modeling (on the formal semantic level) *every* mechanism of encoding information in NL manifested in the structure of NL-expressions of the following kinds:

- texts with direct and indirect speech;
- texts containing compound designations of goals formed from the infinitives or gerunds with dependent words by means of the conjunctions "and," "or" and the particle "not" (such texts may express commands, advices, wishes, obligations, commitments);
- texts containing compound designations of notions;
- texts containing compound designations of sets;
- discourses with references to the meanings of phrases and larger parts of a discourse;
- texts with the word "a notion" ("This notion is used in chemistry and biology," etc.).

That is why it seems that the demands of practice, in particular, the demand to have formal means being convenient for building semantic annotations of arbitrary Web-documents with NL-components show that it is necessary to continue the studies aimed at the elaboration of formal systems of semantic representations with the expressive power being very close to the expressive power of natural language.

One of the possible broadly applicable approaches to this problem is the principal subject of Chaps. 2, 3, 4, 5, and 6.

**Problems**

1. What are the principal restrictions of first-order predicate logic from the standpoint of building mathematical models useful for the designers of semantics-oriented natural language processing systems?
2. Explain the term "the Cartesian product of the sets $X$ and $Y$" without mathematical designations, continuing the phrase "The Cartesian product of the sets $X$ and $Y$ is the set consisting of."
3. What is the difference between the structure of the models of type 1 and type 2?
4. What proposed kinds of models could be of use for the design of NL-interfaces to (a) recommender systems, (b) autonomous intelligent robots?
5. What kinds of the proposed models could be of use for the design of machine translation systems?
6. How does the structure of the models of type 5 reflect the fact that the information for formulating an answer to a request posed by an end user of an intelligent full-text database can be accumulated step by step, as a result of analyzing not one but several informational sources?
7. Why can the partial models of types 1–5 contribute to increasing the quality of documentation of semantics-oriented NLPSs?
8. What are the main principles of Cognitive Linguistics concerning the study of natural language comprehension?
9. Explain the term "an underspecified semantic representation of a NL-text."
10. What is the purpose of introducing the episodic operator?