Chapter 8
A New Method of Transforming Texts into Semantic Representations

Abstract This chapter sets forth a new method of describing the transformation of an NL-text (a statement, a command, or a question) into its semantic representation. According to this method, the transformation includes three phases: (a) Phase 1: The component-morphological analysis of the text; (b) Phase 2: The construction of a matrix semantic-syntactic representation (MSSR); (3) Phase 3: The assembly of a semantic representation of the text, proceeding from its MSSR.

8.1 A Component-Morphological Representation of an NL-text

Let's agree that in this and the next chapters we will consider as lexical units (or word forms) not only separate words but also compound verbal forms ("has been received" and so on), compound prepositions, compound terms ("Olympic games," "artificial intelligence" and so on). This approach allows for attracting the attention to central problems of developing the algorithms of semantic-syntactic analysis by means of abstracting from the details of text preprocessing (it is reasonable to consider such details at the level of program implementation).

Let's say that *elementary meaningful units of texts* are all lexical units, the constructs (the designations of the values of different numeric parameters: 780 km, 12 kg, 7 percent, and so on), the markers (punctuation marks), and the expressions in quotes or apostrophes – the names of various objects.

In order to determine the notion of a matrix semantic-syntactic representation (MSSR) of an NL-text, we will introduce a number of additional data structures associated with the input texts of applied intelligent systems with respect to a considered linguistic basis.

8.1.1 Morphological Representation

Temporarily skipping a number of mathematical details, we'll suppose that *a morphological representation* of a text T with the length nt is a two-dimensional array

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_8,

Rm with the names of columns *base* and *morph*, where the elements of the array rows are interpreted in the following way.

Let *nmr* be the number of the rows in the array *Rm* that was constructed for the text T, and k be the number of a row in the array *Rm*, i.e. $1 \leq k \leq nmr$. Then $Rm[k, base]$ is the basic lexical unit (the lexeme) corresponding to the word in the position p from the text T. Under the same assumptions, $Rm[k, morph]$ is a sequence of the collections of the values of morphological characteristics (or features) corresponding to the word in the position p.

Definition 8.1. Let $Tform$ be a text-forming system of the form (7.4), $Morphbs$ be a morphological basis of the form (7.3), $T \in Texts(Tform)$, nt be the length of the text T. Then *a morphological representation of the text* T is such two-dimensional array *Rm* with the indices of the columns *base* and *morph* that the following conditions are satisfied:

1. Every row of the array *Rm* contains information about a certain word from the input text T, i.e. if *nmr* is the number of the rows in the array *Rm*, then for each i from 1 to *nmr*, such position p can be found in the text T, where $1 \leq p \leq nt$, that
$$t_p \in W,\ Rm[i, base] = lcs(t_p),$$
$$Rm[i, morph] = fmorph(t_p).$$
2. For each word from the text T, there is a row in *Rm* representing morphological information about this word, i.e., for each position p in the text T, where $1 \leq p \leq nt$, $t_p \in W$, there is such k, where $1 \leq k \leq nmr$, that
$$lcs(t_p) = Rm[k, base],$$
$$fmorph(t_p) = Rm[k, morph].$$
3. Every two rows in the array *Rm* differ either due to different basic lexical units in the column *base* or due to different collections of the values of morphological properties in the column with the index *morph*. This means that if
$$1 \leq k \leq nmr,$$
$$1 \leq q \leq nmr,\ k \neq q,$$
then either $Rm[k, base] \neq Rm[q, base]$
or $Rm[k, morph] \neq Rm[q, morph]$.

Thus, any row from *Rm* points a basic lexical form and a collection of the values of morphological properties connected with a certain lexical unit from the text T. At the same time, for each lexical unit from T, a corresponding row can be found in *Rm*.

Example 1. Let T1 be the question “What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi-agent systems (9) ‘Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents’ (10) by (11) professor (12) Fomichov (13) ?

(14).” The text T1 is marked-up in the following way: every elementary expression from the text is followed by the number of elementary meaningful unit of text including this expression. Then a morphological representation *Rm* of the text T1 may have the following form:

base	morph
what	md_1
Russian	md_2
publishing house	md_3
released	md_4
in	md_5
the work	md_6
on	md_7
multi-agent systems	md_8
by	md_9
Professor	md_{10}
Fomichov	md_{11}

Here $md_1, \ldots, md_{11}$ are the numerical codes of morphological features collections, that is connected in corresponding the words from input the text T1. In particular, md_3 encodes the next information: the part of speech – noun, the subclass of part of speech – common noun, the number – singular, the case – common.

The collection of non-negative integers md_{11} encodes the following information: the part of speech – noun, the subclass of part of speech – proper noun, the number – singular, the case – common.

8.1.2 Classifying Representation

Let $Tform$ be a text-forming system of the form (7.4),

$$T \in Texts(Tform),\ nt = length(T).$$

Then, from an informal point of view, we will say that *a classifying representation of the text T coordinated with the morphological representation Rm* of the text T, is a two-dimensional array *Rc* with the number of the rows *nt* and the column with the indices *unit*, *tclass*, *subclass*, *mcoord*, in which its elements are interpreted in the following way.

Let k be the number of any row in the array *Rc* i.e. $1 \leq k \leq nt$. Then $Rc[k, unit]$ is one of elementary meaningful units of the text T, i.e. if $T = t_1 \ldots, t_{nt}$, then $Rc[k, unit] = t_k$.

If $Rc[k, unit]$ is a word, then $Rc[k, tclass]$, $Rc[k, subclass]$, $Rc[k, mcoord]$ are correspondingly part of speech, subclass of part of speech, a sequence of the collections of morphological features’ values.

If $Rc[k, unit]$ is a construct (i.e. a value of a parameter), then $Rc[k, tclass]$ is the string *constr*, $Rc[k, subclass]$ is a designation of a subclass of informational unit that corresponds to this construct, $Rc[k, mcoord] = 0$.

Example 2. Let T1 be the question "What Russian publishing house released in the year 2007 the work on multi-agent systems 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' by professor Fomichov?" Then a classifying representation *Rc* of the text $T1$ coordinated with the morphological representation *Rm* of $T1$ may have the following form:

unit	tclass	subclass	mcoord
What	pronoun	nil	1
Russian	adject	nil	2
publishing house	noun	common-noun	3
released	verb	verb-in-indic-mood	4
in	prep	nil	5
the year 2007	constr	nil	0
the work	noun	common-noun	6
on	prep	nil	7
multi agent systems	noun	common-noun	8
book-title	name	nil	0
by	prep	nil	9
Professor	noun	common-noun	10
Fomichov	noun	proper-noun	11
?	marker	nil	0

Here the element *book-title* is the name of the monograph "Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents."

Definition 8.2. Let $Tform$ be a text-forming system of the form (7.4), $Morphbs$ be a morphological basis of the form (7.3), $nt = length(T)$, $T \in Texts(Tform)$, *Rm* be a morphological representation of T. Then *a classifying representation of the text T coordinated with Rm* is a two-dimensional array *Rc* with the indices of the columns *unit*, *tclass*, *subclass*, *mcoord*, and the number of the rows *nt*, satisfying the following conditions:

1. For $k = 1, \ldots, nt$, $Rc[k, unit] = tk$.
2. If $1 \leq k \leq nt$ and $t_k \in W$, then

$$Rc[k, tclass] = prt(t_k),\ Rc[k, subclass] = subprt(t_k),$$

and it is possible to find such q, where $1 \leq q \leq nrm$, *nmr* is the number of the rows in *Rm*, that

$$Rc[k, mcoord] = q,\ Rm[q, base] = lcs(t_k),$$

$$Rm[q, morph] = fmorph(t_k).$$

3. If $1 \leq k \leq nt$ and $t_k \in Constr$, then

$$Rc[k, tclass] = constr,\ Rc[k, subclass] = tp(infconstr(t_k)),$$

$$Rc[k, mcoord] = 0.$$

4. If $1 \leq k \leq t$ and $t_k \in Names(T form)$, then

$$Rc[k, tclass] = name,\ Rc[k, subclass] = nil,$$

$$Rc[k, mcoord] = 0.$$

5. If $1 \leq k \leq nt$ and $t_k \in Markers$, then

$$Rc[k, tclass] = marker,\ Rc[k, subclass] = nil,$$

$$Rc[k, mcoord] = 0.$$

Thus, a classifying representation of the text T sets the following information:

1. For each lexical unit, it indicates a part of speech, a subclass of part of speech (if it is defined), and the number of a row from the morphological representation *Rm* containing the numerical codes of morphological characteristics corresponding to this lexical unit.
2. For each construct it indicates the class *constr* and a subclass, that is the sort of information unit corresponding to this construct.
3. For each element from the set $Names(T form)$, it indicates the class *name*, the subclass *nil*, and the number 0 in the column *mcoord*.
4. For each separator (the punctuation marks), it indicates the class *marker*, the subclass *nil*, and 0 in the column *mcoord*.

Definition 8.3. Let $T form$ be a text-forming system of the form (7.4), $T \in Texts(T form)$. Then *a component-morphological representation (CMR) of the text T* is an ordered pair of the form

$$(Rm, Rc),$$

where *Rm* is a morphological representation of the text T, *Rc* is a classifying representation of the text T coordinated with *Rm*.

8.2 The Projections of the Components of a Linguistic Basis on the Input Text

Let *Lingb* be a linguistic basis of the form (7.9), and *Dic* be one of the following components of *Lingb* : the lexico-semantic dictionary *Lsdic*, the dictionary of verbal – prepositional semantic-syntactic frames $V fr$, the dictionary of prepositional semantic-syntactic frames Frp. Then *the projection of the dictionary Dic on the input text* $T \in Texts(T form)$ is a two-dimensional array whose rows represent all data from *Dic* linked with the lexical units from T.

Let's introduce the following denotations to be used in this and next chapters:

- $Arls$ is *the projection of the lexico-semantic dictionary Lsdic on the input text* $T \in Texts(Tform)$;
- $Arvfr$ is *the projection of the dictionary of verbal – prepositional frames* Vfr *on the input text* $T \in Texts(Tform)$;
- $Arfrp$ is *the projection of the dictionary of prepositional frames* Frp *on the input text* $T \in Texts(Tform)$.

Example 1. Let T1 be the question "What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi-agent systems (9) 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' (10) by (11) professor (12) Fomichov (13) ? (14)" Then the array $Arls$ may have the following form:

ord	sem	st1	st2	st3	comment
2	Country(z1, Russia)	space.ob	nil	nil	nil
3	publish-house	org	ints	space.ob	nil
4	releasing1	sit	nil	nil	comment1
4	releasing2	sit	nil	nil	comment2
7	work1	sit	nil	nil	comment3
7	work2	inf.ob	dyn.phys.ob	nil	comment4
9	sem1	field-of-activ	nil	nil	comment5
12	sem2	ints	dyn.phys.ob	nil	nil
13	sem3	ints	dyn.phys.ob	nil	nil

where

$$sem1 = multi\_agent\_systems,$$

$$sem2 = certn\,person * (Qualif, professor),$$

$$sem3 = certn\,person * (Surname, \text{“}Fomichov\text{”}),$$

$comment1$ = "This film was released in 2005,"

$comment2$ = "Yves released her hand,"

$comment3$ = "This work took 3 h,"

$comment4$ = "This work was sent via DHL,"

$comment5$ = "a scientific – technical field of studies."

The elements of the column *ord* (ordered number) are the ordered numbers of the rows from the classifying representation Rc, that is, the ordered numbers of elementary meaningful units (or tokens) of the text T.

The number of the rows of the array $Arls$ corresponding to one elementary meaningful lexical unit (i.e., corresponding to one row of the classifying representation Rc) is equal to the number of different meanings of this lexical unit.

The purpose of considering a two-dimensional array $Arvfr$ is as follows: for each verbal form from the text T, this array contains all templates (in other terms, frames) from the dictionary Vfr enabling a linguistic processor to find the possible conceptual (or semantic) relations between a meaning of this verbal form and a meaning of a word or word group depending on this verbal form in a sentence from the text T.

Example 2. Let T1 be the question "What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi agent systems (9) 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' (10) by (11) professor (12) Fomichov (13) ? (14)" Then a fragment of the array $Arvfr$ may have the following form:

nb	semsit	fm	refl	vc	trole	sprep	grc	str	expl
4	releasing1	ftm	nrf	actv	Agent2	nil	1	org	expl1
4	releasing1	ftm	nrf	passv	Agent2	by	1	org	expl2
4	releasing1	ftm	nrf	actv	Product1	nil	1	inf.ob	expl3
4	releasing2	ftm	nrf	actv	Agent1	nil	1	ints	expl4
4	releasing2	ftm	nrf	actv	Object1	nil	1	dyn.phys.ob	expl5

Here 4 is the position of the verb "released" in the considered text T1; ftm (form with time) is the indicator of the verbs in indicative and subjunctive mood; nrf is the indicator of non reflexive verbs; $actv$ and $passv$ are the values "active" and "passive" of the voice; $Agent1$, $Agent2$, $Product1$, $Object1$ are the designations of thematic roles, 1 is the numeric code of the common grammatical case in English; org, $inf.ob$, $ints$, $dyn.phys.ob$ are the sorts "organization," "informational object," "intelligent system," "dynamic physical object,"

$expl1$ = "The studio released (this film in 2005),"

$expl2$ = "(This film) was released by the studio (in 2005),"

$expl3$ = "(The studio) released this film (in 2005),"

$expl4$ = "Yves released (her hand for several seconds),"

$expl5$ = "Yves (released her hand) for several seconds,"

where the auxiliary parts of the examples are surrounded by brackets.

The connection of the array $Arvfr$ with the array $Arls$ is realized by means of the column $semsit$. A template (frame) from the array $Arvfr$ being the m-th row of $Arvfr$ is associated with the row k of the array $Arls \Leftrightarrow$ when this template and the row k correspond to the same lexical unit from the text, and

$$Arls[k, sem] = Arvfr[m, semsit].$$

In the same way the array $Arfrp$ can be built, it is called the projection of the dictionary of prepositional frames Frp on the input text. This array is intended for representing all data from the dictionary Frp relating to the prepositions from the text T and to the empty preposition nil (in case the text T contains the word combinations of the form "Noun1 + Noun2").

Example 3. Let T1 be the question "What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi-agent systems (9) 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' (10) by (11) professor (12) Fomichov (13) ? (14)" Then a fragment of the array $Arfrp$ may have the following form:

prep	sr1	sr2	grc	rel	ex
on	inf.ob	field-of-activ	1	Field1	"a book on art"
on	phys.ob	phys.ob	1	Location1	"a house on the hill"

8.3 Matrix Semantic-Syntactic Representations of NL-Texts

Let's consider a new data structure called *a matrix semantic-syntactic representation (MSSR) of a natural language input text T*. This data structure will be used for representing the intermediate results of semantic-syntactic analysis on an NL-text.

An MSSR of an NL-text T is a string-numerical matrix $Matr$ with the indices of columns or the groups of columns

$$locunit, nval, prep, posdir, reldir, mark, qt, nattr,$$

it is used for discovering the conceptual (or semantic) relations between the meanings of the fragments of the text T, proceeding from the information about linguistically correct, short, word combinations. Besides, an MSSR of an NL-text allows for selecting one among several possible meanings of an elementary lexical unit.

The number of the rows of the matrix $Matr$ equals nt – the number of the rows in the classifying representation Rc, i.e., it equals the number of elementary meaningful text units in T.

Let's suppose that k is the number of arbitrary row from MSSR $Matr$. Then the element $Matr[k, locunit]$, i.e., the element on the intersection of the row k and the column with the index $locunit$, is *the least number* of a row from the array $Arls$ (it is the projection of the lexico-semantic dictionary $Lsdic$ on the input text T) corresponding to the elementary meaningful lexical unit $Rc[k, unit]$.

It is possible to say that the value $Matr[k, locunit]$ for the k-th elementary meaningful lexical unit from T is *the coordinate of the entry to* the array $Arls$ corresponding to this lexical unit.

The column $nval$ of $Matr$ is used as follows. If k is the ordered number of arbitrary row in Rc and $Matr$ corresponding to an elementary meaningful lexical unit, then the initial value of $Matr[k, nval]$ is equal to the quantity of all rows from $Arls$ corresponding to this lexical unit, that is, corresponding to different meanings of this lexical unit.

When the construction of $Matr$ is finished, the situation is to be different for all lexical units with several possible meanings: for each row of $Matr$ with the ordered number k corresponding to a lexical unit, $Matr[k, nval] = 1$, because a certain meaning was selected for each elementary meaningful lexical unit.

For each row of $Matr$ with the ordered number k associated with a noun or an adjective, the element in the column $prep$ (preposition) specifies the preposition (possibly, the void, or empty, preposition nil) relating to the lexical unit corresponding to the k-th row.

Let's consider the purpose of introducing the column group

$$posdir(posdir_1, posdir_2, \ldots, posdir_n),$$

where n is a constant between 1 and 10 depending on program implementation. Let $1 \leq d \leq n$. Then we will use the designation $Matr[k, posdir, d]$ for an element located at the intersection of the k-th row and the d-th column in the group $posdir$.

If $1 \leq k \leq nt$, $1 \leq d \leq n$, then $Matr[k, posdir, d] = m$, where m is either 0 or the ordered number of the d-th lexical unit wd from the input text T, where wd governs the text unit with the ordered number k.

There are no governing lexical units for the verbs in the principal clauses of the sentences, that is why for the row with the ordered number m associated with a verb, $Matr[m, posdir, d] = 0$ for any d from 1 to n.

Let's agree that the nouns govern the adjectives as well as govern the designations of the numbers (e.g. "5 scientific articles"), cardinal numerals, and ordinal numerals.

The group of the columns *reldir* consists of semantic relations whose existence is reflected in the columns of the group *posdir*. For filling in these columns, the templates (or frames) from the arrays $Arls$, $Arvfr$, $Arfrp$ are to be used (the method can be grasped from the analysis of the algorithm of constructing a matrix semantic – syntactic representation of an input NL-text stated in the next chapter).

The column with the index *mark* is to be used for storing the variables denoting the different entities mentioned in the input text (including the events indicated by verbs, participles, gerunds, and verbal nouns).

The column *qt* (quantity) equals either zero or the designation of the number situated in the text before a noun and connected to a noun.

The column *nattr* (number of attributes) equals either zero or the quantity of adjectives related to a noun presented by the k-th row, if we suppose that $Rc[k, unit]$ is a noun.

Example. Let T1 be the question "What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi-agent systems (9) 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' (10) by (11) professor (12) Fomichov (13) ? (14)." The morphological and classifying representations *Rm* and *Rc* of T1, the possible projections of the dictionaries $Lsdic$, Vfr, Frp on the input text T1 are considered in two preceding sections.

With respect to the arrays *Rm*, *Rc*, $Lsdic$, Vfr, Frp constructed for the text T1, its MSSR can have the following form:

locunit	nval	prep	posdir	reldir	mark	qt	nattr
0	1	nil	0, 0	nil, nil	nil	0	0
1	1	nil	3, 0	conc1, nil	nil	0	1
2	1	nil	4, 0	Agent2, nil	x1	0	1
3	1	nil	0, 0	nil, nil	e1	0	0
0	1	in	0, 0	nil, nil	nil	0	0
0	1	in	4, 0	Time, nil	nil	0	0
6	1	nil	4, 0	Product1, nil	x2	0	0
0	1	on	0, 0	nil, nil	nil	0	0
7	1	on	7, 0	Field1, nil	nil	0	0
0	1	0	7, 0	Name1, nil	x2	0	0
0	1	by	0, 0	nil, nil	nil	0	0
8	1	by	7, 0	conc2, nil	x3	0	0
9	1	by	7, 0	conc3, nil	x3	0	0
0	1	nil	0, 0	nil, nil	nil	0	0

where

$$conc1 = Country(z1, Russia),$$

$$conc2 = Qualif(z1, Professor),$$

$$conc3 = Surname(z1, \text{“}Fomichov\text{”}).$$

The constructed matrix reflects the final configuration of the MSSR $Matr$. It means that all semantic relations between the text units were found.

8.4 A New Method of Transforming NL-Texts into Semantic Representations

The concepts introduced above and the stated principles provide the possibility to formulate a new method of transforming the NL-texts (in particular, the requests, statements, or commands) into semantic representations (SRs) of texts.

8.4.1 Formulation of the Method

The proposed method is intended for designing the dialogue systems and includes the following three stages of transformation:

Transformation 1: A component-morphological analysis of the input text.

The essence of the first transformation is as follows. Proceeding from an NL-text T, one constructs one or several component-morphological representations (CMR) of the text T. This means that one constructs one or several pairs of the form (Rm, Rc), where Rm is a morphological representation of the text, i.e., a representation of possible values of the morphological properties for the components of the text T being lexical units (contrary to the numerical values of the properties, to the markers, and to the expressions in apostrophes or inverted commas); Rc is a classifying representation of the text.

In other words, the first transformation consists in (a) distinguishing such fragments of the text (called further the elementary meaningful textual units) that each of these fragments either is a marker (comma, semi-colon, etc.) or an expression in inverted commas or in apostrophes or is associated with certain meaning (or meanings); (b) associating one or several collections of the values of morphological properties (a part of speech, a number, a grammatical case, etc.) with each elementary textual unit being a word or a word group, (c) associating a semantic item (a sort) with each elementary meaningful textual unit belonging to the class of constructs: the numbers and the expressions like “1200 km,” “70 km/h,” etc. For instance, the combination “were delivered” is associated with the part of speech “a verb,” the plural, and the past simple tense.

In the major part of cases, only the CMR will correspond to the separate phrases from the input text. If there are several variants of dividing the input text T into the elementary meaningful units or several parts of speech can be associated with any text unit, then the computer system puts the questions to the end user of the dialogue system, and ambiguities are eliminated after the processing of the answers of the end user to these questions.

Transformation 2: The construction of a matrix semantic-syntactic representation (MSSR) of the text.

The first goal of this transformation is to associate with each elementary textual unit that is not a marker or not an expression in inverted commas or in apostrophes one definite meaning from the collection of several meanings linked with this unit. For instance, the verb "to deliver" has, in particular, the meanings "to deliver a lecture" and "to deliver a thing," and the noun "a box" is linked with two different meanings "a box as a container" and "a box as a theater concept."

The second goal of the Transformation 2 is to establish the conceptual relationships between various elementary textual items and, in some cases, between larger items (e.g., between the meaning of a noun and the meaning of an attributive clause).

Since it is done step by step, the MSSR initially is underspecified. In order to eliminate the ambiguities, the system can apply to the end users with diverse questions. Each new step is able to modify the current configuration of the built MSSR as a consequence of obtaining new information from the analysis of a text's fragment and of inscribing this new piece of information into the MSSR.

During this process, mainly the data from the considered linguistic database (LDB) are used and, besides, the knowledge about the admissible manners to combine the various text units into linguistically correct combinations.

Transformation 3: The assembly of a semantic representation of the input NL-text.

The purpose of this transformation is to "assemble" a semantic representation (SR) of the considered text T, proceeding from the information stored in its MSSR. It is important to note that such SR of T is an expression of an SK-language. That is, it is a K-representation of the text T.

An algorithm transforming an MSSR $Matr$ of an input NL-text T into a formal expression $Semrepr \in Ls(B)$, where B is the conceptual basis being the first component of the considered marked-up conceptual basis Cb, and $Ls(B)$ is the SK-language in the basis B, will be called *an algorithm of semantic assembly*.

Example 1. Let T1 be the question "What (1) Russian (2) publishing (3) house (3) released (4) in (5) the (6) year (6) 2007 (6) the (7) work (7) on (8) multi-agent systems (9) 'Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents' (10) by (11) professor (12) Fomichov (13) ? (14)" The morphological and classifying representations Rm and Rc of T1, the possible projections of the dictionaries $Lsdic$, Vfr, Frp on the input text T1, and a matrix semantic–syntactic representation $Matr$ of T1 are considered in three preceding sections.

With respect to the previous examples concerning the analysis of the question T1, its possible K-representation *Semrepr*1 can be as follows:

$$Question(x1, Situation(e1, releasing1 * (Time, certn\,mom*$$

$$(Earlier, \#now\#) : t1)(Agent2, certn\,publish - house * (Country,$$

$$Russia) : x1)(Product1, certn\,work2 * (Field1, multi - agent\,systems)$$

$$(Name1, Title1)(Authors, certn\,person * (Qualif, professor)$$

$$(Surname, \text{“}Fomichov\text{”}) : x3) : x2))),$$

where *Title*1 is the string “Mathematical Foundations of Representing the Content of Messages Sent by Computer Intelligent Agents.”

Figure 8.1 illustrates the proposed method of transforming NL-texts into their semantic representations.

8.4.2 The Principles of Selecting the Form of a Text's Semantic Representation

The form of a semantic representation of an NL-text T to be assembled from the data stored in the MSSR *Matr*, in the classifying representation *Rc*, and in the two-dimensional array *Arls* – the projection of the lexico-semantic dictionary *Lsdic* on the input text T is to depend on the kind of T.

Let's consider the examples illustrating the recommendations concerning the choice of the form of an SR being an expression of a certain SK-language, that is, being a K-representation of T. In these examples, the SR of the input text T will be the value of the string variable *Semrepr* (semantic representation).

Example 2. Let T1 = “Professor Igor Novikov teaches in Tomsk.” Then

$$Semrepr = Situation\,(e1, teaching * (Time, \#now\#)$$

$$(Agent1, certn\,person * (Qualification, professor)$$

$$(Name, 'Igor')\,(Surname, 'Novikov') : x2)$$

$$(Place1, certn\,city * (Name, 'Tomsk') : x3)).$$

Example 3. Let T2 = “Deliver a box with details to the warehouse 3.” Then

$$Semrepr = (Command\,(\#Operator\#, \#Executor\#, \#now\#, e1)$$

$$\wedge Target\,(e1, delivery1 * (Object1, certn\,box1 *$$

$$(Content1, certn\,set * (Qual - compos, detail)) : x1)$$

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

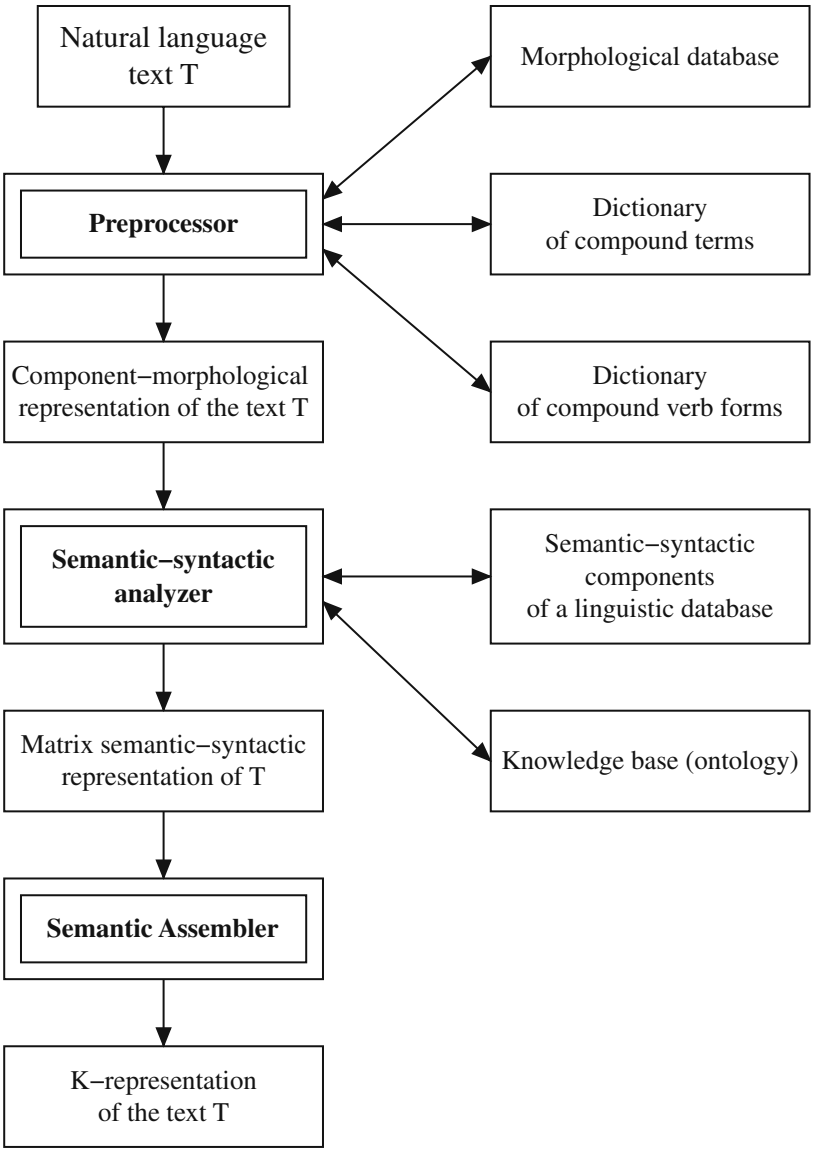

Fig. 8.1 The scheme of transforming NL-texts into their K-representations

Example 4. Let T3 = "Did the international scientific conference 'COLING' take place in Asia?" Then

$$Semrepr = Question\,(x1,\,(x1 \equiv$$

$$Truth - value\,(Situation\,(e1,\,taking\_place*$$

$$(Time, certn\,moment * (Earlier, \#now\#) : t1)$$

$$(Event1, certn\,conference * (Type1, international)$$

$$(Type2, scientific)\,(Name, 'COLING') : x2)$$

$$(Place, certn\,continent * (Name, 'Asia') : x3))))).$$

Example 5. Let T4 = "What publishing house has released the novel 'Winds of Africa'?" Then

$$Semrepr = Question\,(x1, Situation\,(e1, releasing1 *$$

$$(Time, certn\,moment * (Earlier, \#now\#) : t1)$$

$$(Agent2, certn\,publ - house : x1)$$

$$(Product1, certn\,novel1 * (Name1, 'Winds\,of\,Africa') : x2))).$$

Example 6. Let T5 = "What foreign publishing houses the writer Igor Somov is collaborating with?" Then

$$Semrepr = Question\,(S1, (Qual - compos\,(S1, publish - house *$$

$$(Type - geographic, foreign)) \wedge$$

$$Description(arbitrary\,publish - house * (Element, S1) : y1,$$

$$Situation\,(e1, collaboration * (Time, \#now\#)$$

$$(Agent1, certn\,person * (Occupation, writer)$$

$$(Name, 'Igor')\,(Surname, 'Somov') : x1)$$

$$(Organization1, y1))))).$$

Example 7. Let T6 = "Who produces the medicine 'Zinnat'?" Then

$$Semrepr = Question\,(x1, Situation\,(e1, production1 *$$

$$(Time, \#now\#)\,(Agent2, x1)$$

$$(Product2, certn\,medicine1 * (Name1, 'Zinnat') : x2))).$$

Example 8. Let T7 = "For whom and where the three-ton aluminum container has been delivered from?"

$$Semrepr = Question\,((x1 \wedge x2),$$

$$Situation\,(e1, delivery2 *$$

$$(Time, certn\,moment * (Earlier, \#now\#) : t1)$$

$$(Recipient, x1)\,(Place1, x2)$$

$$(Object1, certn\,container1 * (Weight, 3/ton)$$

$$(Material, aluminum) : x3))).$$

Example 9. Let T8 = "How many people did participate in the creation of the textbook on statistics?" Then

$$Semrepr = Question\,(x1, ((x1 \equiv Numb(S1))$$

$$\wedge Qual - compos\,(S1, person) \wedge$$

$$Description\,(arbitrary\,person * (Element, S1) : y1,$$

$$Situation\,(e1, participation1 *$$

$$(Time, certn\,moment * (Earlier, \#now\#) : t1)$$

$$(Agent1, y1)\,(Type - of - activity, creation1 *$$

$$(Product1, certn\,textbook * (Area1, statistics) : x2))))).$$

Example 10. Let T9 = "How many times Mr. Stepan Semenov flew to Mexico?" Then

$$Semrepr = Question\,(x1, ((x1 \equiv Numb\,(S1))$$

$$\wedge Qual - compos\,(S1, sit) \wedge$$

$$Description\,(arbitrary\,sit * (Element, S1) : e1,$$

$$Situation\,(e1, flight * (Time, certn\,moment *$$

$$(Earlier, \#now\#) : t1)\,(Agent1, certn\,person *$$

$$(Name, 'Stepan')(Surname, 'Semenov') : x2)$$

$$(Place2, certn\,country * (Name, 'Mexico') : x3))))).$$

Problems

1. What is the difference between the morphological and classifying representations of the input NL-text?
2. What is the connection between the quantity of the rows of the two-dimensional array *Arls* – the projection of the lexico-semantic dictionary *Lsdic* on the input text corresponding to a lexical unit and the quantity of the meanings associated with this lexical unit?
3. How are the following interpreted: (a) the columns *locunit* and *nval*, (b) the groups of columns *posdir* and *reldir* of a matrix semantic-syntactic representation?
4. What new expressive mechanisms of SK-languages are to be used for building K-representations of the questions about the number of objects characterized with the help of the verbs with dependent words?