

# Context-Dependent Mechanisms for Recognition of Semantic Role of a Word in the Sentence

*Olga A. Nevzorova*

Lexical ambiguity is a serious problem in development of the automatic text processing systems. This paper describes the mechanisms allowing in some cases to resolve the ambiguity of semantic role of a word in the sentence taking parametric description of the given context of semantic role as a basis. The method is based on main results of linguistic theory of G. Zolotova, developed in order to describe Russian syntax. The key idea is construction of a calculus of Russian sentences' models as a combination of semantic-syntactical elements, named syntaxemes.

Basic descriptions of nominal syntaxemes allowed to describe the structure of formal context for different types of sentences. Recognition of semantic role of a word (syntaxeme) is put into effect on the basis of mechanism correlating formal and real parameters of the context describing the semantic role of a word.

## 1 INTRODUCTION

Over the last years the development of network technologies caused intensive research in text processing. A great number of different tasks exists in this field. For our purposes we select annotation systems, specific information extraction systems and acquisition knowledge systems.

The basic problem of text analysis system is an interpretation of the text, i.e. the reflection of text components into certain given structures, which shows the type of relations between components of the text. Hereafter, we will discuss the analysis (interpretation) of a separate sentence (not discourse), i.e. construction of the reflection of components in the sentence into domain structures. In order to interpret the text properly, a problem of recognition of semantic role of the sentence's interpreted component (word or segment) must be solved. This problem is often solved in different ways.

We distinguish the following approaches to solution of the given problem. Among the most interesting works are Cardie's [Cardie 1996], in which a lazy-learning approach for both (morpho)syntactic and semantic disambiguation is proposed. Ng and Lee [Ng and Lee 1996] report results of statistical methods usage in word sense disambiguation. Other approaches

are connected with usage of various restrictions on sublanguage and on domain of application.

In this article certain mechanisms of choosing the word's semantic role are described. These mechanisms allow to resolve the ambiguity of a word's semantic role using parametric context of semantic role.

## 2 LINGUISTIC BASIS OF THE WORK

The main results described in this article are achieved when developing the Russian technical texts processing. However, as it will be shown below, the main ideas of our approach don't depend on the realization language. Our purpose is to show the constructive character of the approach and to discuss the main problems of its application.

Linguistic theory of G. Zolotova underlays our approach. Its results allow to describe relations between semantic and syntax from uniform positions and to develop clear criterions of building the hierarchic system for different types of russian sentences. Linguistic results are systematized in the dictionary [Zolotova 1989]. Let us describe briefly the main G.Zolotova's results, for we need them to understand this article.

Over contemporary development of linguistics the most actual problem is classification and distinguishing elementary syntax units of which other, more complicated units, are built.

The system that solved this problem for Russian syntax was proposed by Zolotova G. Key concept of syntax in Zolotova's opinion is the concept of the syntaxeme. Syntaxeme is simple and indivisible semantic-syntactical unit of Russian language serving as well as the elementary meaning carrier and as the constructive component of more complicated syntactical units, which are characterized by definite set of syntax functions.

Language gives us an opportunity to express our minds by means of syntax rules by predicative conjunction of syntaxemes. A sentence of any composition from the point of view of its organization and dismemberment can be considered to be one of the combinations of syntaxemes.

The idea of the component analysis of a sentence is certainly not new, but it is the first time this problem has been solved for the model of Russian syntax. Another key idea is the development of the Case grammar conception in the direction of research into proper syntactical, constructive qualities of case forms. Russian grammar consists of different cases and case-preposition forms with different functions, that's why functional-syntactical aspect of the given problem is particularly actual for Russian grammar.

So, each syntaxeme is characterized by three features - form, meaning and function. We distinguish the case and case-preposition forms of the

syntaxemes. The meaning of the syntaxeme is its case role in D.Fillmore's sense. The essential characteristics of the syntaxeme is its syntactical function which determines its ability to realize itself syntactically under certain positions of the sentence. We distinguish three main functions of syntaxemes:

1. Independent and isolated usage.
2. Usage as the component of a sentence.
3. Usage as the component of collocation.

Function 1 is usually used in titles. The concept of functions 2 and 3 is detailed by more specific concept of position. That means that the function is realized in certain syntactical position in any type of sentence or collocation model.

Consequently, the problem of syntaxeme's functional type determination requires development of the sentence's models' formal calculus. For Russian grammar this problem was put and successfully solved in above-mentioned work. Let us enumerate positions of distinguished functions.

The following positions are distinguished in function 2:

- (1) position of a Component bearing predicative sign and executing a role of grammatical subject in the sentence;
- (2) position of Component - predicative sign, executing a role of grammatical predicate in the sentence;
- (3) position of a Component - situative spreader of the sentence's model;
- (4) position of a Component - half-predicative complicator of the sentence's model.

The following positions are distinguished in function 3:

- (1) Component of a verb phrase;
- (2) Component of a noun phrase;
- (3) Component of an adverb or adjective phrase .

We distinguish three following functional types of a syntaxeme:

1. Free syntaxeme (can be used in functions 1,2,3).
2. Stipulated syntaxeme (can be used in functions 2,3, less often in 1).
3. Connected syntaxeme (can be used only in function 3).

The meaning of free syntaxemes (locative, temporative, directive, deliberative and others) doesn't depend on their positional opportunities. The meaning of connected and stipulated syntaxemes (subjective and objective) is derived from the position they realize. Let us demonstrate one explanatory example. Locatives are considered to be free syntaxemes. They can take almost any position in any of three functions.

1. In headlines: *U vorot* “**At the gate**”
2. As the component of the sentence:
  - 2.1 as the grammatical subject: *U vorot ljudno*  
“*There are many people at the gate*”.  
  
(In English, I think, locative can't take the grammatical subject position).
  - 2.2 as the grammatical predicate: *Deti u vorot*  
“*There are the children at the gate*”.  
  
(In English, I think, locative can't take the grammatical predicate position).
  - 2.3 as the complicator of the sentence's model: *On dognal nas u vorot*  
“*He has caught us up at the gate. (He has caught us when we were at the gate)*”.
  - 2.4 as the spreader of the sentence's model: *U vorot laet sobaka*  
“*The dog barks at the gate*”.
3. Attached to the word spreader:
  - 3.1 Attached to the verb: *sidet' u vorot*  
“*to sit near the gate*”.
  - 3.2 Attached to a substantive: *skamja u vorot*  
“*the bench at the gate*”.
  - 3.3 Attached to an adjective: *gustye u vorot (kusty)*  
“*thick at the gate (bushes)*”.

Application of stipulated syntaxemes is restricted by limits of the sentence. Outside the sentence they do not have correlative meaning. For example syntaxemes with categorial meaning of the person realize in dative case the function of subject carrying predicative sign of state or age:

*Stariku ne zdorovitsja. Rebenku god*  
“*The old man is not in good health. The child is one year old*”.

These syntaxemes carry out the only function number 2. They are not used in attached to the word position (3). In headlines the homonym with the meaning of a person is used, but it has another function - it is free syntaxeme of the addresser:

*Dorogomu uchitelju*  
“*To dear teacher*”.

Functional classification of syntaxemes gives us the answer to basic questions of sentence's theory. We learn about predicative minimum of the sentence; about opportunities and ways to express main components of different models of the Russian sentence; about opportunities and ways to spread models; about character and degree of attached to word connection.

One of the most important results of this theory and realization of syntaxemes' potential combinative opportunities is logical description of Russian sentence's model. Constructed calculus of models confirms a thesis of principal two-composition of Russian sentence. That means, predicative minimum of the Russian sentence consists of two main components, that express an idea of correlation between predicate (from the point of view of tense, person and modality) and object - the carrier of predicative sign. But grammatical means of predicative sign and object expression may be different in various models of the sentence. Certain regular means for grammatical subject expression in different types of Russian sentences can be distinguished. Those are stipulated syntaxemes with the meaning of personal subject of predicative sign (*On v trevoge* "He is anxious"; *U nego zar* "He has fever heat"; *Emu veselo* "He is enjoying himself"), free syntaxemes with locative meaning, some temporative ones in sentences of place and time with unpersonal subject: (*Za oknom moroz* (*morozit, morozno*) "It is a frost out of the window"; *Seichas – polnoch* "It is midnight now").

Thus, different case forms of grammatical subject are not obstacles for recognizing the case role of corresponding syntaxemes as subjects carrying predicative sign.

Thereby, the problem of a main component of Russian sentence (personal or traditionally supposed to be unpersonal, without a subject) is consistently but not undoubtedly solved. This problem is not very sharp in other languages, for example in languages with fixed word order and, thus, with fixed number of sentences types. In English grammar not only the formal position of sentence's grammatical subject is fixed, but also its grammatical forms. However, ambiguity of prepositions, which are combined with syntaxemes in sentences, sharpens the problem of nominal syntaxemes' semantic role recognition for these languages. This question will be discussed in the next paragraph.

### 3 NOMINAL SYNTAXEME'S SEMANTIC ROLE RECOGNITION

Main results of Zolotova's theory are presented in Syntactical dictionary [Zolotova 1989].

Hereafter we will examine an organization of its part 1 *Nominal syntaxemes*. The dictionary can be used both for synthesis of right sentences in Russian and analysis of Russian sentences.

Let us precisely define our problem and resources we have at our disposal for its solving. We consider a problem of technical texts' analysis in order to check if the description of information for definite problem field is complete. To estimate completeness of text information, we must build an

interpretation of the text in object structures of the field. Construction of interpretation is connected with recognition of the meaning of distinguished elementary text sentence's units and establishment of correlations between them.

Elementary units of the sentence are segments. The concept of a segment is the generalization of the concept of a syntaxeme. The segment, unlike the syntaxeme, can have (in common case) an internal structure and is characterized by form, meaning and syntactical function. To recognize a segment's meaning the above mentioned Syntactical dictionary is used. However its direct application has some difficulties. The point is that in common case the dictionary determines the syntaxeme's meaning ambiguously by its form and syntactical position in the sentence. The ambiguity of the meaning of a syntaxeme, which has fixed morphological form and syntactical position, is the first problem. This problem is partially solved by indication of semantic class of the word-syntaxeme. The second problem is recognition of syntactical position of the syntaxeme in the sentence. In the worst case the number of syntactical positions of the syntaxeme is 8, usually it is 3 or 4.

The next step of our discussion was to introduce the concept of a context of the syntaxeme's semantic meaning. Thereby, the problem of recognition of the syntaxeme's meaning is defined as a problem of recognition of the syntaxeme's context meaning.

Formal context of the syntaxeme is set by parametric description, which consists of the following parts:

1. Type of the context;
2. Semantic role of the syntaxeme;
3. Syntaxeme's parameters;
4. Parameters of environment;
5. Sentence's parameters;
6. Special methods of the context examination;
7. Parameters of the text;
8. Key parameters.

Detailed description of the formal context's structure follows below:

1. Type of the context (identifier)
2. Semantic role of the syntaxeme:
  - 2.1 Semantic role of the syntaxeme
  - 2.2 Syntactical function of the syntaxeme
  - 2.3 Additional semantic characteristic (context supplement)
3. Parameters of the waited syntaxeme:
  - 3.1 Semantic type
  - 3.2 Morphological form
  - 3.3 Sintagmatical characteristic (preposition, particle *no*)
4. Parameters of environment:
  - 4.1 Syntactical type of environment (borders of the context)
  - 4.2 Characteristics of the main element:

- 4.2.1 Semantic characteristic
- 4.2.2 Morphological form
- 4.3 Fixed parameters of a distinguished word
- 4.4 Fixed words (lists of words)
- 5. Parameters of the sentence:
  - 5.1 Parameters of the predicate:
    - 5.1.1 Presence of the verbal predicate
    - 5.1.2 Semantic type of the predicate
    - 5.1.3 Active (passive) form of the verbal predicate
    - 5.1.4 Positive (negative) form of the verbal predicate
    - 5.1.5 Type of predicative group
  - 5.2 Parameters of the grammatical subject:
    - 5.2.1 Presence of the grammatical subject
    - 5.2.2 Semantic type
    - 5.2.3 Morphological form
  - 5.3 Fixed words (lists of words)
- 6. Methods of the context examination:
  - 6.1 Calculated parameters of the context
- 7. Text characteristics:
  - 7.1 Part of the text (title or paragraph)
- 8. Key parameters

#### 4 RECOGNITION OF THE SYNTAXEME'S SEMANTIC ROLE WITH THE HELP OF THE CONTEXT

Given description of the formal context of the syntaxeme's semantic role is a model of an entry in the new syntaxemes' Context Vocabulary, which is being elaborated on the basis of Syntactical dictionary. Elaborated format of dictionary's entry is maximal description. Specific entries of the dictionary consist of different sets of description parameters. The entry is formed on the basis of unstructured description of the corresponding entry in Syntactical dictionary. As a whole, the format of new vocabulary is a generalized and structured set of description parameters, extracted from text descriptions of Syntactical dictionary entries.

Usage of the formal context for recognition of the syntaxeme's semantic role allows to solve the given problem successfully. Actual context of a sentence is formed when analyzing the sentence. Recognition of the semantic role is realized on the basis of identification mechanism of the formal and actual syntaxeme's contexts. When parameters of the compared contexts coincide, coincidence on the key parameters is especially important. In this situation, the semantic role of corresponding formal context is ascribed to analysed syntaxemes. Gradual analysis of the syntaxemes precises the parameters of the sentence's actual context. As a result conforming description of the context is formed.

## 5 EXAMPLE OF AN ENTRY IN THE CONTEXT DICTIONARY AND ANALYSIS OF THE SPECIFIC SYNTAXEME

Let us describe an example of vocabulary entries setting descriptions of definite contexts of the free syntaxeme in the genitive case without a preposition in the meaning of the genitive date. This syntaxeme can be potentially realized in six syntactical positions. Let us examine the usage of this syntaxeme in two main positions – as the grammatical subject and as the predicate of the sentence:

1. *14 fevralja - den' Svjatogo Valentina* “**February,14** is the Saint Valentine’s Day”. (The position of the grammatical subject)
2. *Prem'era – vtorogo janvarja* “The premiere will take place on **January,2**”. ( The position of the predicate).

Peculiarity of syntactical organization of the examined Russian examples is in the absence of the verbal predicate, though the second sentence can be formed using phasic verbs (*Prem'era sostoitsja 2 janvarja* “The premiere **will take place on January,2**”). But in this case the syntaxeme *January,2* is used in the meaning of sentence’s predicate as the component of the predicate group.

Let us give an example of descriptions of the entries in Context Dictionary for the given syntaxeme in the genitive case with the meaning of genitive date.

1. K24 – identifier of the context
- 2.1. genitive date
- 2.2. II.1
- 3.1. semantic type "date"
- 3.2. genitive case
- 4.1. definition with numeral words
- 4.2.1. name of the month
- 4.2.2. genitive case
- 4.4. names of the months
- 5.1.1 verb is absent
- 5.1.2. qualificatory predicate
- 5.1.5. [link-verb *to be* +] <name> (model of qualification)
- 5.2.2. type "date"
- 5.2.3. genitive case
- 5.3. names of the months
- 6.1. checking the format of the date
8. 3.1.



1. K25 - identifier of the context
- 2.1. genitive date
- 2.2. II.2
- 3.1. semantic type "date"
- 3.2. genitive case
- 4.1. definition with numeral words
- 4.2.1. name of the month
- 4.2.2. genitive case
- 4.4. names of the months
- 5.1.1. absent verb or phasic verbs (to take place, to begin, to last, to proceed)
- 5.1.2. predicate of events' timing {the model "Event-Time"}
- 5.1.5. [phasic verb +] <date> (model of the predicate)
- 6.1. checking the format of the date
8. 3.1.

General scheme of the sentence's analysis includes the following typical components:

- TAGGING of ELEMENTS (level of the sentence) defines the base of elements in a sentence and gathering of statistical data on the elements of the sentence.
- MORPHOLOGICAL TAGGING (level of words) assigns the element-words a set of morphological parameters, using special morphological dictionaries;
- SEGMENTATION (level of the sentence) forms the basic units of the sentence – segments (simple and complex) on the basis of elements.
- SEMANTIC TAGGING (level of segments) recognizes semantic roles of segments basing on semantic roles of syntaxemes as stated in the Context Dictionary.
- ANALYSIS (level of the sentence) displays the segments to structures showing relations between them.

Let us examine the stage of SEMANTIC TAGGING, where the segments' semantic role is being recognized (in the individual case of the nominal syntaxeme). On the input of algorithm is the syntaxeme in definite grammatical form, the semantic role of the syntaxeme generates on the output.

For sentences 1) and 2) on the previous stages of text processing the base of elements with statistical information on the types of elements is formed. Morphological tagging of the element-words is assigned, the base of segments with statistical information on the types of created segments is formed *ibidem*.

For sentences (1) and (2) two segments are distinguished. They are given in brackets:

(1) *[14 fevralja]*- *[den' Svyatogo Valentina]*.

*[February,14]* *[is the Saint Valentine's Day]*.

(2) *[Prem'era]* - *[vtorogo janvarja]*.

*[The premiere]* *[will take place on January,2]*.

The results of SEMANTIC TAGGING of the sentences are:

(1)

*[14 February]* semantic role of the "genitive date" is in the function of grammatical subject;

*[is the Saint Valentine's Day]* has the semantic role of qualificatory predicate. The given segment has the internal structure and consists of two simple segments. One of them is the main segment, which participates in forming the predicative model, another one is playing the role of "genitive possessive".

(2)

*[The premiere]* has the semantic role of "subject of predicate-timing event" in the function of grammatical subject;

*[will take place on January,2]* has the semantic role of the "genitive date" in the function of predicate.

Let us show on the example of sentence 1). how the problem of ambiguous choice of the segment's semantic role is solved. Marked segments of the sentence are analysed gradually from the start of the sentence. Statistics of the segment types which forms individual parameters of the sentences' actual context are used in analysis. Specifically we use the parameter establishing the number of nominal segments and the parameter fixing the absence of verbal segment. Other parameters of the actual context are the parameters of group 3 (semantic type of the nominal segment) and the parameters characterizing the environment of syntaxeme (group 4). Given the meanings of these parameters for the first nominal segment from the Context Dictionary the possible contexts are being selected. The number of possible contexts is 2. Hereafter the parameters of formal and actual context are consequently compared. We determine that the context with identifier K24 attributes the semantic role of "genitive date" in the function of grammatical subject in the model of the sentence with the qualificatory predicate to the first nominal segment

The following purpose is to find the nominal syntaxeme capable to take the position of qualificatory predicate. For this purpose we extract the contexts of the syntaxemes in the nominative case, with the predicate in the qualification models from the Context Dictionary. Such context (K16, brought below) exists and the analysed syntaxeme satisfies its limits. So, the model of the sentence is completely determined.

1. K16 – identifier of the context
- 2.1. qualificatory predicate
- 2.2. II.2
- 3.1. list of semantic types
- 3.2. nominative case
- 4.1. predicate group
- 4.2.2. nominative case
- 5.1.1. verb is absent
- 5.1.2. qualificatory predicate model
- 5.1.5. [link-verb "to be" +] <name> (model of qualification)
- 6.1. hyphen presence check
8. 3.2, 5.1.5

## 6 CONCLUSIONS

Mechanism for comparing the formal and actual contexts allows to identify the model of the sentence and to solve the problem of recognition of the nominal segments' (syntaxemes) semantic roles within the limits of fixed contexts.

Final conclusions on the offered mechanism will be made after the development of full Context Dictionary of the semantic roles of the different types syntaxemes is over. Realization of this current purpose is a complicated task, nevertheless we consider that all preconditions for its eventual success are present.

## REFERENCES

- Zolotova, G.A., 1990. *Syntactical dictionary. Repertoire of the elementary units of the Russian syntax*. M., "Science", 1988. 440 pp. In Russian.
- Ng, H.T. and H.B. Lee, 1996. *Integrating multiple knowledge sources to disambiguate word sense: an exemplar-based approach*. In Proceedings of the annual meeting of the ACL, ACL-96.
- Cardie, C., 1996. *Embedded Machine Learning Systems for Natural Language Processing: A General Framework*. In S. Wermter, E. Riloff, and G. Scheler, (Eds.), *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing*, pp.315-328. Berlin:Springer-Verlag.

**Olga A. Nevzorova** is a senior scientific researcher of Kazan State University, Russia. She can be reached at [Olga.Nevzorova@ksu.ru](mailto:Olga.Nevzorova@ksu.ru), tel. (7-8432) 387491.