# AN EXPERIMENT ON SYNTHESIS OF RUSSIAN PARAMETRIC CONSTRUCTIONS

I.S. Kononenko, E.L. Pershina

AI Laboratory, Computing Center,
Siberian Branch of the USSR Ac.Sci.,
Novosibirsk 630090, USSR

## ABSTRACT

The paper describes an experimental model of syntactic structure generation starting from the limited fragment of semantics that deals with the quantitative values of object parameters. To present the input information the basic semantic units of four types are proposed:"object", "parameter", "function" and "constant". For the syntactic structure representation the system of syntactic components is used that combines the properties of the dependency and constituent systems: the syntactic components corresponding to wordforms and exocentric constituents are introduced and two basic subordinate relations ("actant" and "attributive") are claimed to be necessary. Special attention has been devoted to problems of complex correspondence between the semantic units and lexical-syntactic means, In the process of synthesis such sections of the model as the lexicon, the syntactic structure generation rules, the set of syntactic restrictions and morphological operators are utilized to generate the considerable enough subset of Russian parametric constructions.

## 1 INTRODUCTION

The semantics of Russian parametric constructions deals with the quantitative values of object parameters. The parametric information is more or less easily explicated by means of basic semantic units of four types: "object" ('table', 'boy'), "parameter" ('weight', 'length', 'age'), "function" ('more', 'equal', 'almost equal') and "constant" ('two meters', 'from 3 to 5 years').

In simple situations each of these units is separately realized in a lexeme or a phrase, their combinations forming full expressions with the given sense: malchik vesit bolshe dvadcati kilogrammov 'boy weights more than twenty kilograms'. It is precisely these direct and simple means of expressions that are usually used in systems generating natural language texts.

Natural languages, however, operate with more complex means of expression ; one-to-one correspondence between semantic units and lexical items is not always the case. The complex situations are suggested here to be explained in terms of decomposition of the input semantic representation (cf. the notion of form-reduction in Bergelson and Kibrik (1980)). This phenomenon is exemplified by such Russian lexemes as stometrovka 'hundred-meters-long-distance' which semantically incorporates the four constituents of the parametric semantics.

As an ideal, a language model should embrace mechanisms that provide generation and understanding of the constructions that make use of the various possibilities of lexicalization and grammaticalization of sense. The presented model deals with some aspects of the phenomena that have not been considered before: all the possibilities of decomposition of the input information are taken into account and the means of syntactic structure representation are developed to provide the synthesis of the parametric syntactic structure.

The paper is organized as follows. In section 2 the set of semantic components is described. In section 3 the relevant syntactic notions are introduced. In section 4 the process of synthesis is outlined, followed by conclusions in section 5.

## 2 SEMANTIC COMPONENTS

The information to-be-communicated is represented as a set of four semantic units each of them being marked with the type-symbol (o - "object", p - "parameter", f - "function", c - "constant").

At the initial step of synthesis a process involving the decomposition of the input semantic structure into a system of semantic components takes place. Usually, a semantic structure corresponds to several decompositions. The forming of a component may be motivated by the following reasons.

In the event of separate lexicalization a component represents exactly one semantic unit. There are four components of this kind according to the number of unit types. So, the object component $K_o$ represents a unit of the "object" type and is realized in a noun (dom 'house') or a possessive adjective (papin 'father's'). The parameter component $K_p$ is lexicalized in parametric nouns, verbs and participles. The function component $K_f$ is realized in lexemes of different syntactic classes: prepositions, comparative verbs and participles and forms of comparative degree of some adjectives and adverbs. The constant component $K_c$ corresponds to measure adjectives and some quantitative constructions described in Kononenko et al. (1980).

A component represents more than one semantic unit in two situations.

(1) The first one has been mentioned above. It concerns the phenomenon of incorporation of several units in one lexeme: thus, the component $K_{opfc}$ is introduced to account for the lexemes like stometrovka and $K_{pf}$ component is a prototype of parametric-comparative adverbs like shire 'wider'.

(2) On the other hand, the introduction of a component may be connected with the fact that a certain unit is not lexicalized at all. Such "reduced" elements of sense are considered to be realized on the surface by the type of the syntactic structure composed of the lexicalized units of the component. For example, in Russian approximative constructions litrov pjat 'about-five-liters' it is only the "constant" unit that is lexicalized and the unit of the "function" type ('almost equal') is expressed by purely syntactic means, i.e. the inverted word-order in the quantitative phrase. The corresponding component represents both the "function" and "constant" units.

## 3 SYNTACTIC STRUCTURES

The syntactic structures of Russian parametric constructions are various enough. The full system of rules (Kononenko and Pershina, 1982) provides the generation of nominal phrases and simple sentences but the structures within the complex sentence such as komnata, dlina kotorojj ravna pjati metram 'room whose length is five meters' are left out of account. So, the model allows for the following examples: shestiletnijj malchik 'six-years-old boy'; bashnja vysotojj bolee sta metrov 'tower of more than hundred meters height'; kniga stoit pjat rublejj 'book costs five roubles' etc.

To represent the syntactic structures the system of syntactic components suggested in Narinyani (1978) proved to be useful, that combines the properties of the dependency and constituent systems. Two different types of syntactic components, the elementary and non-elementary ones, are claimed to be necessary. **The elementary component corresponds to a wordform and is traditionally represented by a lexeme symbol marked with syntactic and morphological features.**

**The non-elementary component is composed of syntactically related elementary components.** The outer syntactic relations of the non-elementary component cannot be described in terms of syntactic and morphological characteristics of the constituent elementary components. The notion of a non-elementary component is a convenient tool for describing the syntactic behaviour of Russian quantitative constructions composed of a noun and a numeral: the morphological features of the subject quantitative phrase (nominative, plural) are not equivalent to those of the nominal constituent (genitive, singular).

The minimal syntactic structure that is not equal to a wordform is described in terms of a syntagm, i.e. a bipartite pattern in which syntactic components are connected by an actant or attributive syntactic relation. Each component is marked with the relevant syntactic and morphological features.

The actant relation holds within the pattern in which the predicate component X governs the form of the actant component Y, e.g.: shirina [X] ehkrana [Y] 'width of-screen' the governing lexeme shirina determines the genitive of the noun-actant.

The attributive relation connects the component X with its syntactic modifier, or attribute, Y. The attributive syntagm is typically composed of a noun and an adjective (stometrovaja [Y] vysota [X] 'one-hundred-meters height'), a noun and a participle, a noun and another noun, a verb and an adverb or a preposition.

The syntactic relation is represented by an "act" or "attr" arrow leading from X to Y.

The syntactic class features reflect the combinatorial properties of the components in the constructions under consideration. The following are some examples of the syntactic features:

"$S_{obj}$" - object nouns (dom 'house')

"$S_{param}$" - parametric nouns (ves 'weight')
"$A_{poss}$" - possessive adjectives (papin 'father's')
"$V_{param}$" - parametric verbs (stoit 'to-cost')
"$P_{param}$" - parametric participles (vesjashhijj 'weighing')
"$A_{meas}$" - measure adjectives (pjatiletnijj 'five-years-old')

The syntactic structure does not contain any syntactically motivated morphological features connected with government or agreement (the latter are described separately in the morphological operators section of the model). The case of the noun used as attribute is reflected in the syntactic structure representation since this feature is relevant in distinguishing syntagms.

## 4 STRUCTURE GENERATION

The first step of synthesis is the decomposition of the input semantic representation into the set of semantic components. The possibilities of lexicalization of components are determined by the lexicon that provides every lexeme with its semantic prototype - the set of semantic units incorporated in the meaning of the lexeme. The lexicalization rules replace the semantic components by the concrete lexemes, e.g.:'weight' $[K_p]$ is replaced by the lexemes ves $[S_{param}]$, vesit $[V_{param}]$ or vesjashhijj $[P_{param}]$.

The semantic types of components determine their combinatorial properties on the syntactic level. The grammar is developed as the set of rules each of which provides all the syntagms realizing the initial pair of components.

For example, the pair $\{K_o, K_p\}$ corresponds to six syntagms:

(a) $A_{poss} \xleftarrow{attr} S_{param}$ papin ves 'father's weight'

(b) $S_{obj} \xrightarrow{attr} S_{param,gen}$ ehkran shiriny 'screen of-width (gen)'

(c) $S_{obj} \xrightarrow{attr} S_{param,instr}$ bashnja vysotojj 'tower of height (instr.)'

(d) $S_{obj} \xrightarrow{attr} P_{param}$ kniga stojashhaja 'book costing'

(e) $S_{obj} \xleftarrow{act} V_{param}$ malchik vesit 'boy weights'

(f) $S_{obj} \xleftarrow{act} S_{param}$ vysota doma 'height of-house'

The rules applicable to different fragments of the same decomposition are bound with the syntagmatic restrictions that prevent the unacceptable combinations of syntagms. Thus the combination of the syntagm (c) for $\{K_o, K_p\}$ and the adjective lexicalization of the "constant" component forms the unacceptable syntactic structure *ehkran pjatimetrovojj shirinojj 'screen of 5-meters-long width (instr)'.

The process of synthesis yields all the possible syntactic structures corresponding to the input semantic representation.

## 5 CONCLUSION

**In this report on the basis of the very limited data of the parametric constructions an attempt has been made to consider a simplified model of synthesis of the text expression beginning from the given semantic representation. The scheme presented above is planned to be implemented within the framework of the question-answering system.**

Right from the start of synthesis the process of decomposition of the input semantics takes place in order to capture different cases of complex correspondence between the semantic units and the lexical -syntactic means. To generate the considerable enough subset of Russian parametric constructions such sections of the language model as the lexicon, the grammar generating the syntactic structures, the set of syntactic restictions and morphological operators are utilized. The listed constituents, however, do not, exhaust all the necessary mechanisms of synthesis since the problems of word-order are left to be investigated and an additional reference to various aspects of the communicative setting is required. We believe that being of primary importance for automatic synthesis of natural language texts the communicative aspect of text generation presents one of the most promising research directions for future activity.

## 6 REFERENCES


Bergelson, M.B.; Kibrik, A.E., 1980. "Towards the General Theory of Language Reduction". In: Formal Description of Natural Language Structure. pp. 147-161. Novosibirsk (in Russian).

Kononenko, I.S.; Krasnova, V.A.; Pershina, E.L., 1980. The Structure of Russian Quantitative Constructions. Preprint No. 237. Novosibirsk (in Russian).

Kononenko, I.S.; Pershina, E.L., 1982. A Model Generating Syntactic Structures of Some Russian Parametric Constructions. In: Formal Representation of Linguistic Information. pp. 103-122. Novosibirsk (in Russian).

Narinyani, A.S. 1978. Formal Model: General Scheme and Choice of Adequate Means. Preprint No. 107. Novosibirsk (in Russian).