

## ON THE TYPES OF SEMANTIC COMPRESSION OF TEXT

N. N. Perceva

Faculty of Numerical Mathematics and Cybernetics  
Moscow State University, U.S.S.R.

Any text is constructed in accordance with implicit principles of semantic compression. The paper attempts to describe some of these principles.

There are two main types of text semantic compression: (1) the omission of "meaningful" parts, and (2) the omission of standard connections among parts of the text.

1. The first type of compression is subdivided into: (a) omission within a predication (by a predication is meant the structure: "predicate - P, its actants - A"), and (b) omission of the whole predication.

a. In special conditions a part of a predication should be understood as the whole predication in accordance with 6 basic rules<sup>+/</sup>:  $A \rightarrow AP$ ,  $P \rightarrow AP$ ,  $A...P \rightarrow AP$ ,  $A_1...A_j \rightarrow$

$\rightarrow A_1PA_j$ ,  $A_1 \rightarrow A_1PA_j$ ,  $A_1 \rightarrow A_j$ .

Examples.

$P \rightarrow AP$

The patient was operated.  $\rightarrow$  The patient was operated by surgeons.

$A_1...A_j \rightarrow A_1PA_j$

After the operation Nick feels good.  $\rightarrow$  The operation was made on Nick.

-----  
<sup>+/</sup> In general, as it seems, rules of a text grammar, unlike most rules of most sentence grammars, should be not obligatory, but optional.

b. The omission of a predication is diagnosed by the presence of another predication ( $PR_j$ ), which is connected with the omitted predication (PR) by such relations (j) as "part" (including starting, intermediate or final point or interval), "attribute", "accompanying", "reaction", "cause", etc. (the number of such relations seems not to exceed 20). In such cases the understanding of the text includes the rule:

$PR_j \rightarrow PR.$

Examples.

j = attribute

to put the foot in the stirrup  $\rightarrow$  to mount a horse

j = reaction

The Sahalinians take measures to liquidate the consequences of the typhoon. The roads and bridges are being reconstructed. -- The roads and bridges were destroyed by the typhoon.

2. The omission of standard connections among the parts of the text can either be universal or depend on the type of the text. The texts can be divided into two main groups: "physically oriented" and "mentally oriented". Roughly speaking, physically oriented texts describe events as they occur and group the events in accordance with their sequence, while mentally oriented texts describe the events as they are reflected in the text author's mind and group the events into (individual or socially acknowledged) classes.

An important subclass of (2) constitute the cases of "determining". We call a determiner a unit which relates not only to its immediate context, but to some larger piece of the text, up to the text as the whole. (Determiners can be either manifested or omitted in the text.) Let us compare two texts:

- (i) Yesterday Nick went to the theatre. Today he is going to the cinema.
- (ii) Yesterday Nick went to the theatre. He met his friends. On returning home they discussed the performance.

In (i) the meaning 'yesterday' relates only to one sentence, while in (ii) this meaning relates to the whole text. A piece of the text connected with a definite determiner is called its "sphere of operation". Along with constant determiners (such as one in text (ii)) there are variable determiners: their meaning is regularly changed while moving through the text (something like a counter). An example of a variable determiner is a sequence of predicate time markers in physically oriented texts, which usually are partially omitted and should be reconstructed during the text analysis.

To process a constant determiner is to find its sphere of operation, to process a variable determiner is, besides, to find out a proportional change of its meaning.

The reconstruction of text omissions demands using different kinds of information, among which an important role belongs to "common-sense" knowledge. This knowledge is divided into general (typical relations of objects and events) and concrete. The concrete knowledge is necessary, for instance, to understand such a sentence as Flamingoes bones exploration shows that the age of the birds is about 50 million years, where the linguistic meaning of the word age contradicts the concrete knowledge of the life limits of creatures, which causes the rereading of the sentence.

Another interesting type of information needed is the information about a possibility of a "subjective" description (from the point of view of either the author or the object described).

Examples.

Through the bus window she looked at a moving building.

-> It was the bus but not the building which moved.

The people of the Hungarian village of Ekshu are proud of an interesting building - the church shaping a mushroom. -> The church is situated in or near the village of Ekshu.