1967 INTERNATIONAL CONFERENCE

ON COMPUTATIONAL LINGUISTICS

An INTEGRATED LANGUAGE THEORY

S. Abraham

7; 13 Decembrie , Timisoara IV, Rumania

Rumanian Mathematical Society

ABSTRACT. The outlines of an INTEGRATED LANGUAGE THEORY are drawn, with grammar, semantics and text-theory as components. Problems of machine retrieval and machine translation are discussed on the basis of the outlined theory.

At present it is unanimously agreed that no computational linguistics can be developed without an (before - hand constructed) exact language theory. (The computer - handling of such an exact theory is in no way a transparent problem). The aim of the present paper is to outline such an exact integrated language theory.

The most promising basis for such a theory is the generative view on language. Unfortunately the generative theory of language is at present unsatisfactorily developed. First, this is true for generative grammars. But all other language theory constructions rely heavily on grammar (Katz & Fodor, Katz & Postal, Abraham & Kiefer).

At present generative grammar consists of two components : a phrase - structure component and a transformational one (I do not touch upon the problems of the phonemic and morphophonemic levels). In Chomsky's formulation the phrase - structure component necessarily contains context-restricted rules. But this has the unpleasant consequence that the most important decision problems are unsolvable (in this component). Besides this, no satisfactory solution is formulated for the generation of discontinous structures. Chomsky's last formulation of this component has even some more inconveniences. The transformational component has no exact formulation as the transformational rules are unsatisfactorily defined. The problem of analysis is also not satisfactorily solved.

For these reasons I propose a generative grammar also of two components : a matrix component and a transformational component. The matrix component consists of a <u>regular elementary matrix grammar</u>, containing only (ordered) <u>elementary context - free rules</u> , with solvable main decision problems, and which generates in a natural way discontinous structures. Within the transformational component the notion of transformational rule is exactly defined. The problem of an adequate analysis (of the generated sentences), the main problem of generative language theory, is satisfactorily solved by introducing the following analyses (graphs) : morphological (which usually is called the (phrase - structure) derivational graph), syntactical (not considered in Chomsky's variant), configurational (not considered in Chomsky's variant).

At present no (integrated) language theory can be conceived without a semantic component. In the theory which is outlined the semantic component is constructed as follows. A finite number of semantic categories (of the given language) are considered. Each word of the language is characterized by a proper semantic matrix and its definition . The dictionary of the language consists of triplets (word, proper semantic matrix, definition) called lexical issues. Each word of a configuration is characterized by its semantic matrix and its definition. On the basis of the semantic matrices the notion of semantic regularity of configurations is defined. The semantic regularity of a sentence is defined as the condition of the semantic regularity of all the configurations in the configurational analysis of the sentence. The sense of a(semantic regular) sentence is its configurational graph, with all the words substituted by their lexical issues. The notion of truth is introduced (in accordance with Tarski), and three types of analiticity are defined : grammatical, semantical and deductive analiticity.

On the basis of the above outlined theory an exact text component is developed, i.e. a component of the integrated language theory which deals with units larger than the sentence.

The outlined integrated language theory differs essentially from the constructions of Chomsky, Katz & Fodor, Katz et Postal, Abraham & Kiefer.

The exactly (and formally) constructed language theory permits to give a satisfactory solution to the theoretically based machine translation and machine retrieval. This last is formulated in the terms of A -oriented abstract (of a paper), ∅ - oriented abstract, N - oriented abstract, minimal N - oriented abstract, maximal N -oriented abstract and E - maximal superior N - oriented abstract.

BIBLIOGRAPHIE

1 - R. C. CROS, J. C. GARDIN, F. LEVY : "L'automatisation des recherches documentaires. Un modèle général : le SYNTOL." Gauthier-Villars, Paris, 1964, 260 p.

2 - A. BORILLO, M. COYAUD, J. VIRBEL : "Etude sur l'indexation automatique". Convention de recherche nº 65 - FR - 160 de la Délégation Générale à la Recherche Scientifique et Technique. Rapport semestriel nº 1, juin 1966, 37 p.

3 - M. COYAUD : "Une méthode sémantique d'indexation automatique de documents", International Conference on Computational Linguistics, New-York, mai 1965, 32 p.

4 - M. COYAUD et N. SIOT-DECAUVILLE, "L'analyse automatique des documents", Mouton, Paris 1967 (sous presse).

5 - "Un système général de documentation automatique, le "SYNTOL", Presses académiques européennes, Bruxelles, 1964, 4 tomes, 588 p.