XIII.   MECHANICAL TRANSLATION*

V. H. Yngve	G. H. Harman	J. D. McCawley
Carol M. Bosche	Renate I. Heil	T. More, Jr.
Elinor K. Charney	Phyllis Hirshfang	W. K. Percival
Ann Congleton	Muriel S. Kannel	A. C. Satterthwait
J. L. Darlington	E. S. Klima	Rosemarie Sträussnigg
D. A. Dinneen	K. C. Knowlton	M. R. Quillian

A.   SENTENCE PARSING WITH A SELF-ORGANIZING HEURISTIC PROGRAM

Self-organizing heuristic search procedures for automatic sentence parsing have been studied.   Experimental results demonstrate that a statistically guided program can produce correct parsings even with a grammar involving extreme overgeneralization, that strategies exist for producing these parsings efficiently, and that these strategies can be developed automatically from a training sequence of sentences and their correct parsings.

A computer program has been developed which learns to parse correctly and efficiently on the basis of a training sequence of sentences and their correct parsings.   Its strategy of search for the grammatical structure of a sentence is based upon a set of treelike patterns representing local configurations of nodes and words in correctly parsed sentences.   Two numbers are associated with each pattern,   one showing how many times the indicated configuration has been observed in correct parsings, and the other showing how many times the configuration has been constructed during search. The ratio of the first number to the second is taken as the estimated probability, according to the pattern,   that the connection resulting in the indicated configuration is correct.   At any point in the search tree, the strongest such probability estimate determines the next connection for the partial structure corresponding to that point; hence it determines the next move from that point in the tree.   The strategy used for determining the point from which to proceed is always to generate that partial structure with the highest average of probability estimates for all of its connections.   This strategy is in effect a method of repeated reconsideration as to whether the last move made leads toward the right answer, or whether an alternate move from some other point in the search tree may be better.

The parsing program,   coded in the COMIT language for the IBM 7090 digital computer, was used in an experiment with 300 sentences of Basic English.   The machine attempted to parse each of these sentences on the basis of experience gained from all previous attempts and their hand-parsed correct answers. The grammar used could have allowed a very large number of incorrect parsings, because of the overgeneralizations implied by its rules, but only the first parsing found was taken as the machine's answer.

The results indicate that a simple grammar can be used effectively for parsing, that an efficient search procedure can be based upon a set of treelike patterns and associated parameters, and that such a set of patterns and parameters can *be* developed automatically from a training sequence of sentences and parsings.   In the experiment with 300 sentences, few nonsensical parsings resulted;   an incorrect parsing almost always represented either an ambiguity of English which is very difficult to resolve on the basis of syntax even with a "good" grammar,   or a difficulty which could have been resolved by a simple mechanized procedure that does not have to fight the battle of a long combinatorial search.   Even better performance is to be expected from a slightly modified search strategy and from an extended set of patterns;   modifications are also suggested for the self-organization process.

The results of the study suggest a design for a practical parsing mechanism.   It consists essentially of: (a) a linear array of patterns ordered according to the estimated probability that they do the right thing if they apply to a partial construction that is right thus far, and (b) a strategy for selecting from the already produced partial structures one that will probably lead to the correct parsing because the next step of the construction maintains a low average "information content" per connection.

Techniques similar to those developed may be useful in other structure-building areas in which problems consist of completing structures in "typical" ways,   and for which the number of possible structures is so large that useful statistics must be statistics not on structures but on their properties.

<div align="right">K. C. Knowlton</div>