[*International Conference on Machine Translation of Languages and Applied Language Analysis*, National Physical Laboratory, Teddington, UK, 5-8 September 1961]

REPORT ON:

"CONSTRUCTION OF A TEXTUAL ANALYSIS ALGORITHM WITH THE AID OF A COMPUTING MACHINE"

by

OLGA S. KULAGINA
(Steklov Mathematical Institute,
Moscow, U.S.S.R.)

MISS KULAGINA described methods for the determination of syntactic analysis algorithms by a computer. A text was prepared by a linguist, to show the government relations between its words. This text was examined by the computer, and, with the aid of the given government relations, configuration tables were produced. Then, forgetting, as it were, the governments previously given, the computer analysed the text using the configuration tables. It compared its results with the human ones, and pointed out sentences it had erroneously analysed.

This work was done with texts of 500 words in French, English, German and Russian. Four variants of the program were used, scanning the text to the left or the right, and looking for the governing word of a given word, either first on its right, then on its left, or first on its left, then on its right. In the 500 word tests, about 40 errors were made. There was not much fluctuation in this number of errors, but the variant which gave best results was the one which went from left to right through the text, and looked for the governing word, first on the right and then on the left of the governed one.

Miss Kulagina then described complementary rules which could be added; these were called algorithms of the second kind. One example was the discovery of words which were forbidden from occurring between the governing and the governed words. Another sort of complementary rule deals with the sentence

'function f has a derivative'

and prevents 'f' becoming attached to the verb, since it actually is linked to the substantive 'function'. Thus, the complementary rules help to correct errors made in algorithms of the first kind. The 40 errors mentioned above are in this way reduced to 6.

DISCUSSION ON MISS KULAGINA'S PAPER

DR. GARVIN raised the question of human errors made in providing the analysed text to the machine. Could one be sure that all the errors found were due to the machine?

MISS KULAGINA replied that in their experience this was so, they were machine errors.

DR. ZARECHNAK described the difficulties he had experienced in establishing word relationships by computer. Adjacent words which, by themselves, had every appearance of being linked, could in fact be unrelated.

MISS KULAGINA said they hoped that more complicated algorithms would be developed to deal with these difficulties.

In answer to M. CORBE, MISS KULAGINA confirmed that the work was done on a computer, not hand-simulated.

Some further details of the method were elicited in discussions with PROF. JOSSELSON, PROF. OETTINGER and DR. HAYS.

Combinations of only two words (governor and governed) were considered, though more than two words could be involved in the complementary rules. The text fed to the computer was marked with these government relations; for example

From these markings, a table of allowed configurations was built up. In using this table to analyse texts, only one analysis was obtained, though later they hoped to derive all possible analyses. The tables of configurations formed, in effect, the discovered grammar of the language.

D. W. DAVIES