FTD-HT-66-343

FOREIGN TECHNOLOGY DIVISION

UNIVERSAL TRANSLATOR

by

I. Makrushenko

Distribution of this document is unlimited.

UNEDITED ROUGH DRAFT TRANSLATION

UNIVERSAL TRANSLATOR

By: I. Makrushenko

English pages: 4

SOURCE: Leningradskaya Pravda, No. 163, July 12, 1964, p. 4.

UNIVERSAL TRANSLATOR

I. Makrushenko

The front along a female line of a four legged mammal was subjected to the effect of action of our wheeled crew within the framework of a religious ceremony. This is not an abracadabra. And so, if you believe in the Polish fantasist Stefan Lem, who brought an apparatus from a far off planet Aldebaran colorful expressions of the drunken peasant Franek. In the Russian variant his words sounded differently: "an s.o.b. baptized by a Pole". Franek did not understand the too free translation and broke up this extravagently looking apparatus. And to bad, it was capable of translating at once from 196,000 unknown languages.

But perhaps the existence of such a universal translator is possible? Scientists consider it – as possible.

At the Languages Institute of the Academy of Sciences USSR, is a small group of linguists, which is guided by Associate Professor of the Leningrad University, Candidate of Philological Sciences N. D. Andreyev. Their task is – to investigate all language laws, in order to develop a general theory of decoding speech. They were called upon to solve not just partial problems of a certain single language, but to derive general rules suitable for any language, assuming they are

not entirely unknown. They started to create a universal method of investigating all languages with the aid of electron machines, is being derived method to show how easy it would be to decode even "dead" languages, and languages of the fogginess of Andromeda, if such a language ever existed. ( )

There are two accurate mathematical methods suitable for the investigation of human speech. One – probability – allows us to calculate the number of repetitions of language units: such a percentage of single letter words, and such as – 15 letters. The second method – theoretical – multiple – is based on properties of letters to combine with each other. For example, in the beginning of the word after "p" there is often "r" (spring), but it is impossible to find a Russian word beginning with "pshch" or "lr".

Andreyev made the suggestion to combine both methods and to dump the entire necessary calculation work on the shoulders of cybernetic machines. His "key" for translation, requires no knowledge of the sense of the text, nor a dictionary, nor grammar. Needed are only thousands of pages and a machine with a great memory. By purely formal signs, the electronic brain will at first discover the grammar of the language, and then determine the structure of words, and in the final end, their sensual connections.

At first the scientist, working either by hand or on a machine, began determining the frequency of appearance of various letters. In Russian language, the letter "i" has always repeated itself at the end of the word. They inserted an enormous listing into the machine, and set up a program: to determine with what previous letter in these words, "i" is often encountered. It was found to be "o". In the "memory" bag of the machine, words themselves, are not found connected by any grammar: "demoy", "stroy", "novoy", "Zelenoy" and others.

In the entire selection, they then cut off the ending "oy", and began checking with what other endings "dom", "str", "nov", and "zelen" will be encountered. The machine separated a new champion — — with endings "ogo", but in this case threw out words from the listing which began with "dom". You will find in no Russian text, a word "domogo".

This job was being carried out by scientists, who prohibited themselves to penetrate into the sense of words, just as if they did not at all understand the Russian language. They had to investigate to what results the sorting machine leads. And the machine without any reason, hauls in two memory bag words entirely different by grammar: "strogo" and "novogo". The first one — dialect, the second — adjective in genetive case. In suggestions, they behave entirely different. It then appears, that the machine will form no grammar rules?!

The following combination with the ending "uyu", again brought nothing. With the ill-ated "str", the memory bag fell substantively with the accusative case "stroyu", again no grammar!

The stump "str" did not turn out in any of the tests, except with the ending "ykh". These word endings: "novykh", "domov", "zelenykh", "lawns" and "mokrykh" were often encountered, but the word "strykh" was not encountered.

The machine stopped and took up only the words which behave identically and other endings are not encountered. And so, sorting words alternately by the sign of encounters with various endings, or by the sign of repetition of individual letter combinations; the scientists separated a group of words which are similar to each other from the viewpoint, mathematics and linguistics. They were found to be adjectives of a hard declination of "novyy" type. And so, the first morphological type was found separated on the basis of statical-combination modeling

of the Russian language.

The Andreyev group also works on the modeling of English, German, French and Bulgarian; a total of 15 languages. A collection will soon be published summing up three years of the groups work.

Any speech, whether Russian, Ukrainian or ... Marsian; is subject to these general rules – says Andreyev. They can be detected by a universal machine translator with our "key". Of course, this will only be a formal analysis of the unknown language. And the true sense should always be discovered by the human being.

It is not necessary to think that the knowledge of rules of formulating speech will be suitable only when encountering Marsians. In the Andreyev method is a greater terrestrial application. His "key" will find application in enormous motion picture storage houses. Speaking, that it is sometimes easier again to carry out an experiment, than to find a publication about it. In the world every year, are published about 3,000,000 journal reports up to 200,000 patents, about 50,000 scientific books not speaking already about newspaper materials. How can one work in this sea of information?

To aid comes the machine. But first, to code the report is necessary to bring its content into machine language, to teach the machine to convert and prepare thousands of pages of new publications ... without mathematical linguistics, without perfect analysis of terrestrial language, without their models nobody can come through.