

THE COMPUTATION LABORATORY

Harvard University

Cambridge Massachusetts

A TRIP TO THE SOVIET UNION

August 1958

Anthony G. Oettinger

In Moscow. there are four groups working on mechanical translation. All are interested in the notion of an intermediate language. Three are attempting to devise a synthetic language, while the remaining group is planning to use Russian, The groups are at the following institutions:

1. The Linguistics Institute (Institut Jazykoznanija),
2. The Steklov Mathematical Institute (Mathematicheskij Institut imena Steklova),
3. The Electrical Modeling Laboratory (Laboratorija Ehlektromodelorovanija). This laboratory is a branch of Viniti, the All-Union Institute of Scientific and Technical Information.
4. The Institute of Precision Mechanics and Computer Engineering (Institut Točnoj Mexaniki i Vychislitel'noj Tekhniki)

At Leningrad, N.D. Andreev is head of the recently created Experimental Laboratory of Machine Translation (Ehksperimentalnaja Laboratija Mashinnogo Perevoda)¹ and is said to have as many as fifty collaborators from time to time most of whom are students or faculty people primarily engaged in other pursuits. Andreev's group was responsible for much of the exotic language work reported at the May 1953 conference. The work there started in February of 1958 and was interrupted by summer vacation. It is expected that Leningrad will have an Ural machine soon, which Andreev may be able to use.

The Linguistic Institute,. I met the following people from the Linguistics Institute:

1. I. A. Mel'chuk
- 2, L. N, Iordanskaja
- 3, R. M. Frumkina

1. see "Voprosy Jazykoznanija", 1958, No.3.

Mel'chuk told me he is interested in the formulation of an abstract, synthetic intermediary language, not merely as a potential tool for mechanical translation, but also as a way of getting insight into the structure of languages. His system includes an analysis of syntactic configurations and relations: to discover the fundamental configurations and relations thereby expressed. The end product is the expression of this fundamental material in the synthetic intermediate language. The analysis is to proceed by the use of a general method based on a set of basic operators, the analysis of different languages being chosen by setting appropriate parameters in the general method. The expression of the original material in the synthetic language is equivalent to a theory of linguistic structure, adaptable; to all language amenable to analysis.

The analysis is being made on the basis of the examination of one page of Vinogradov's Theory of Numbers in every language that is under consideration. In the case of Armenian it was necessary first to translate this page of Vinogradov into Armenian.

Frumkina is interested in language statistics, with a view toward finding the best words to include in a dictionary.

A. A. Reformat'skij and O. S. Axmanova at this institute, also take an interest in mechanical translation. Professor Reformat'skij is the head of the Applied Linguistics Section of the Institute The Steklov Mathematical Institute. Ljapunov is in charge, and the group consists of the following:

1. O. S, Kulagina
2. G. V. Chekova (now Vokulovskaja)

3. G. Bagranovskaja
4. T. Gavrilova
5. T. N. Moloshnaja

The group has the occasional use of a STRELA computer at the Computing Center of the Academy of Sciences, which is under the direction of Academician A. A. Dorodnitsyn. I was unable to meet this group, since they were away at the time of my trip.

Kulagina is said to be working on a set of twenty fundamental operations for use in an analysis scheme to be applied to the Vinogradov text mentioned on page 2. These are at the register (word) and character levels and include comparison for identity, etc. It is planned also to develop blocks, e.g. homonymy-fixing blocks, since presumably the problems of homograph resolution are quite similar for all languages under study. Kulagina is also said to be working on the mechanization of algorithmization,,

The Electrical Modeling Laboratory. The group here is under the administration of V. A. Uspenskij and includes the following:

1. E. V. Paducheva
2. A. L. Shumilina.
3. Z. M. Volotskaja
4. I. Shalimova
5. M.M. Langleben

V. V. Ivanov, a member of the Editorial Board of the journal "Voprosy Jazykoznanija" is a consultant to the group working on information retrieval problems at the Electrical Modeling Laboratory, and also consults with this group.

I met with Uspenskij, Paducheva, Shumilina, and Ivanov at the Linguistics Institute. Mel'chuk, Iordanskaja, and Frumkina of the Linguistics Institute were also present. Paducheva showed me her scheme for analyzing word-pairs. She has found a set of 43 word pairs, selected as describing the important (she hopes all) relationships that can obtain between two words. Each word in a pair is described by five characteristics. A big problem is to pick these so that a set of descriptors describes a unique relation, and that each word pair maps onto only one descriptor. The words in a pair may come from anywhere within a sentence. Two word pairs, of the forty-three which are distinguished, are the following:

sistema - vektorov

delenie - okruzhnosti

The difference between these two lies in that "delenie" is a verbal derivative, and it is felt that it must be translated by an "-ing" form without intervening "of", while the translation of the first will have an intermediate "of". Having isolated the forty-three types, they now are making lists of word pairs, each identified by their five descriptors, and are cataloguing relations obtaining between all pairs of words in the sentence. In each pair, there is a governing and a governed element.

The Institute of Precision Mechanics and Computer Engineering. A meeting at this institute was arranged for me by V. V. Ivanov. Present were:

1. I. S. Muxin
2. I. K. Bel'skaja
3. T. M. Nikolaeva
4. S. S. Belokrinit'skaja
5. G. A. Tarasova
6. L. N. Bykova
7. V. A. Voronin

Tarasova and Bykova are studying English, Nikolaeva is studying Russian, Belokrinitskaja is studying German, and Voronin is studying Chinese. Japanese is also being studied. A few other full-time staff members were on vacation.

S. A. Lebedev is Director of the Institute, and Muxin is Assistant Director. The Institute has a BESM computer and the group uses this machine.

This group is interested in bidirectional language translation, with Russian as the intermediate language. I was given a copy of T. M. Nikolaeva's "Analiz Russkogo Predlozhenija" (Analysis of the Russian Sentence) a

The work on English at the Institute has included the use of some 3000 sentences (200 passages) of text in hand tests of some of the programs elaborated by the group. Smaller amounts have been treated similarly in the other languages. According to Bel'skaja, the English-Russian dictionary now includes some 3000 English words, of which approximately 600 are polysemantic. Separate algorithms have apparently been devised to resolve the semantic ambiguity of each of these. These polysemantic words are content, rather than function words.

In the field of Chinese, Voronin is said to have devised a means of extracting information about grammatical number by analysis of Chinese sentences.

Panov, who was formerly associated with the Institute, has returned to teaching and is no longer working on mechanical translation.

The Institute of Foreign Languages. I gave a talk about our work to a group of the Institute of Foreign Languages (Pervyj Moskovskij Pedagogicheskiy Institut Inostrannyx Jazykov). "The Bulletin of the Seminar on Problems of Machine Translation" (Bjulletin' Ob"edinenija po Problemam Mashinnogo Perevoda") is published there. Issues No.1 through 7 have been published thus far. While at the Institute, I met V. Ju. Rozentsvejg and I. Revzin. The latter gives a Mathematical Linguistics course there. The main job of the Institute is to train people in foreign languages, and no active research on mechanical translation is going on.

Conclusions:

Soviet work on Automatic Translation appears to be of good quality, and an object of serious effort. Impressions gained from the published literature are consistent with first hand impressions. Lack of machine time is a serious problem for Russian workers, and there is no evidence that they are significantly ahead of other countries, if at all.