

MATHEMATICAL LINGUISTICS AND MACHINE TRANSLATION IN THE
COMPUTING CENTRE OF THE HUNGARIAN ACADEMY OF SCIENCES

by F. Kiefer

Historical

The work on MT from Russian into Hungarian began in 1961 when two teachers of Russian, Gy.Hell and Gy.Sipőczy, made an experiment at the University of Technical Sciences. This experiment did not go, however, beyond a word-by-word translation.

In March 1962 a conference was held jointly by the Computing Centre of the Hungarian Academy of Sciences, the Institute for Linguistics of the Hungarian Academy of Sciences and the Committee for Cybernetics of the Presidency of the Hungarian Academy of Sciences. This conference heard Hungarian research-workers in the field of mathematical linguistics and machine translation give account of their results and discussed some general problems. Gy.Hell and Gy.Sipőczy outlined their algorithms, S.Kónyi pointed out some mistakes in Mel'chuk's algorithm for MT from Hungarian into Russian, Gy.Szépe discussed in a paper some general aspects of MT. Besides the questions of MT, problems of mathematical linguistics were also discussed. Some papers dealt with algebraic linguistics (L.Kalmár, F.Papp and F.Kiefer), some with statistical linguistics (including information theory; A.Rényi, I.Fónagy and others). Naturally, statistical linguistics has an old tradition in Hungary, in the field of algebraic linguistics, however, - as it was documented at the conference - a great gap has to be filled.

In the summer 1962 the MT-group was formed at the Computing Centre within the Department for Theoretical Questions. The group consisted of one full-time, one part-time mathema-

tician and of one part-time linguist. The first task of the group was the programming and checking of the algorithms elaborated by Hell and Sipőczy. The Computing Centre has a computer M-3 which is suitable for experimental purposes but does not make the checking of longer programs possible. In the case of long programs the program must be divided into several parts and checked successively. The other practical difficulty concerns the programming: the writing of programs demands more time than the elaboration of algorithms, in the first place because we have no suitable auto-code-system at our disposal and to write programs in machine language is a time-consuming affair. These circumstances determine to some extent the further research-work of the MT-group, giving it a more theoretical direction.

In September 1962 the Bolyai Society organized a colloquium on the topics "The Foundations of Mathematics, Mathematical Machines and their Applications". A separate session was devoted to mathematical linguistics. On the conference also Y.Bar-Hillel (Israel), R.B.Less (USA), M.Bierwisch (DDR), S.Abraham (Roumania) and P.Sgall (Czechoslovakia) took part and gave lectures on different aspects of mathematical linguistics. As far as Hungarian research-work is concerned, Gy.Hell, L.Dezso and Gy.Sipőczy gave account of their results with respect to the mechanical analysis of Russian texts, F.Kiefer compared Kulagina's set theory model with Chomsky's generative model.

Current research

In September 1963 the MT-group was created a separate unit with an independent research plan with F.Kiefer as the leader of the group. At present the group consists of four full-time mathematicians (plus one guest-mathematician who is spending a year in our Institute) or rather "mathematical-linguists" because three of the mathematicians have also

linguistic qualification) and of two part-time linguists. Besides, a few outside linguists also take part in the work of the group.

The group co-operates with the linguistic group at the University of Debrecen (see below) and with the group for general and structural linguistics of the Institute for Linguistics of the Hungarian Academy of Sciences.

The current research covers the following fields of study.

a.) Machine Translation

First, the algorithms elaborated by Hell and Sipőczy were tested on the computer M-3 and later some parts of them on the computer Elliott-803, too. With respect to some special problems additional algorithms have been elaborated by B.Dömölki and F.Kiefer (dictionary look-up based on a method developed by Cooper, cutting of endings etc.).

All the elaborated algorithms up to the present do not essentially go beyond morphological analysis. The same holds true of the algorithm prepared by D.Varga.

The algorithms worked out by our group so far have all the following faults:

- i.) A recurring problem is that of storage-capacity, which is not a linguistic problem;
- ii.) Each algorithm is influenced by the computer available to us;
- iii.) Some of them are based almost exclusively on the binary system;

- iv.) The algorithms are worked out with high redundancy in that they do not contain essentially new aspects or answers besides solving some purely technical matters;
- v.) The morphological analysis is of less interest at the present state of MT.

Quite naturally then our group has recently turned to the questions of syntactic analysis. At the same time the investigations of semantic problems have come to the fore. For resolving the problems of syntactic analysis partly some purely empirical, partly some methods of model construction are used.

F.Kiefer and S.Abraham are seeking for a general model for MT on the basis of transformational grammar. The Comit-system is used for programming. The first results showed that the translation from English into Rumanian - on the basis of a very restricted corpus, of course - is far easier than the translation from English into Hungarian because of the approximate adequateness and inadequateness of the set of transformations.

A particular group headed by D.Varga has recently tackled some problems of syntactic analysis of Russian on the basis of technical texts. The texts are examined according to different aspects, e.g. agreement-problems, government, word order, participial constructions, verbal constructions etc. The results obtained in this way are important not only from the point of view of MT but also from that of language teaching and the theory of translation.

b.) Algebraic linguistics

F.Kiefer tackled Kulagina's set theoretic model in his first papers and applied it to Hungarian. He found out also the relations between Kulagina's model and Chomsky's transformational grammar and Bar-Hillel's quasi-arithmetical notation, respectively. Some relevant aspects of model construction has re-

cently been pointed out. F.Kiefer and S.Abraham work on a semantic model based on generative grammar. The model assigns to each word a matrix containing semantic categories and also some semantic categories for kernel sentences have been stated.

S.Abraham is also dealing with some of the interconnections of the theory of grammar and the theory of algorithms.

It is, however, still too early to correctly estimate the importance of this work and of the results obtained so far.

c.) Statistical linguistics

In the field of statistical linguistics - at least for the time being - there is no independent research work. As our group co-operates also with the Institute for Linguistics of the Hungarian Academy of Sciences we write also programs for statistical investigations but the interpretation of the results obtained does not come within our scope.

d.) The teaching of mathematical linguistics

F.Kiefer lectures at the Philological Faculty of the University of Budapest for mathematical linguistics for the fourth semester. Mostly students of mathematics and philology attend his lectures as well as teachers and research-workers. The number of attendants is between 15 and 20.

The lecture course has included the following subjects: general aspects of mathematical linguistics, model construction in linguistics, the set theoretic model of Kulagina, the model of Ajdukiewich-Bar-Hillel-Lambek, generative grammar, some further models (predictive analysis and push-down store grammars, projective grammars, dependency grammars etc.), statistical linguistics (above all, the ideas of Herdan and Mandelbrot). Some dissertations are expected in these topics in the near future. From the next semester a particular branch will be called into being at the University for training "mathematical-linguists".

e.) Further works

The MT-group publishes a periodical called "Computational Linguistics" twice a year (chief editor: F.Kiefer, editorial board: F.Papp, S.Petőfi, Gy.Szépe and D.Varga). The periodical gives account of the results of the group and of others being connected with our work. From this issue on foreign research-worker's papers will also be printed in this periodical.

In cooperation with Skriptor (Stockholm), bibliographical service is done at the Centre, i.e. bibliographical cards including all works of mathematical linguistics are edited and sent out to the subscribers of SMIL (Statistical Methods in Linguistics) and to others being on an exchange scheme of publications with our Centre.

F.Kiefer has been appointed Hungarian editor of SMIL which, from the next issue onwards, will deal not only with statistical methods in linguistics but with other methods also e.g. algebraic linguistics etc.

REFERENCES

- KIEFER, F.,
- 1) About a set theoretic model in linguistics, (in Hungarian), in Ált.Nyelv.Tan.I. pp.187-200, Akadémiai Kiadó, Budapest (1963)
 - 2) The Questions of Mathematical Linguistics (in Hungarian), in Valóság 6/1, pp.84-90 (1963)
 - 3) Algorithmus und Programm für das Aufsuchen eines Wortes im Speicher, in CL 1, pp.189-201 (1963)
 - 4) Algorithmus für die Analyse von Endungen, in CL 1, pp.203-221 (1963)
 - 5) Bemerkungen zur Anwendung mengentheoretischer Methoden in der Sprachwissenschaft, in The Foundations of Mathematics, Mathematical Machines and Their Applications (L.Kalmár, ed.), forthcoming
 - 6) Set-theoretic and Mathematico-logical Models in Linguistics (in Hungarian), in Ált.Nyelv.Tan.II., pp. , Akadémiai Kiadó Budapest, forthcoming
 - 7) Some Aspects of Model Construction in Linguistics, in SMIL (Stockholm), forthcoming
 - 8) On categorial grammars, (in Hungarian) in Ált.Nyelv.Tan.III., pp. , Akadémiai Kiadó Budapest, forthcoming
- VARGA, D.,
- 1) Morphological Analysis by Help of the Method of Successive Delimitation in CL 1, pp. 223-255 (1963)

- 2) Проблемы осуществлению морфологического анализа в МП, in Научно - техническая информация, ВИНТИ (forthcoming)
- AGÓCS, I., VARGA, D., Взаимосвязь элементов отношения русских и венгерских имён, in CL 2, pp. (1964)
- HELL, GY.,
- 1) Some Theoretic and Practical Aspects of MT from Russian into Hungarian; (in Hungarian), in Ált. Nyelv. Tan. II., (forthcoming)
 - 2) О некоторых характерных чертах алгоритма машинного перевода с русского языка на венгерский, in Slavica II (Debrecen), pp. 151-164 (1962)
 - 3) Определение номинальных групп в МП с русского на венгерский, in CL 1, pp. 5-107, (1963)
 - 4) The two Levels of the Syntactic Analysis in the Mechanical Analysis of Russian Texts, in CL 2, pp. (1964)
- SIPÓCZY, GY.,
- 1) The Programming of Prepositional Constructions (in Hungarian), in Ált. Nyelvt. Tan. II., (forthcoming)
 - 2) Algorithms for the Analysis of Homonymic Nominal Forms, in CL 1, pp. 109-157, (1963)
 - 3) Вопросы программированию русских предложных оборотов, in Slavica III. (Debrecen), pp. 151-169, (1963)
 - 4) Some Semantic Aspects of the MT from Hungarian into Russian, in CL 2, pp. (1964)

ВСТОС, J., Опыт автоматической сегментации текста при МП
с русского языка на венгерский, in CL 2, pp.
(1964)

PETŐFI, S. J., Graphs in Linguistics, (in Hungarian), in Ált.
Nyelv. Tan. III., (forthcoming)