MACHINE LANGUAGE TRANSLATION*

BY

ERWIN REIFLER

I. Historical Outline

In January, 1954, newspapers, radio, and television all over the world reported a new advance in the application of electronic computers. In these reports the World Headquarters of the International Business Machines Corporation of New York announced that a 701-type calculator had been applied to translation work and had actually translated a number of specimen Russian sentences into English. Thus the world at large learned for the first time of the coming new miracle of machine translation or, as it is now generally called, MT.

The preparatory studies for this IBM experiment had been done at the Institute of Languages and Linguistics of Georgetown University in Washington, D. C.[1] Such research had, however, been carried on earlier by others. L. I. Zhirkov, discussing a report by O. S. Kulagina and I. A. Mel'chuk on the linguistic aspects of machine translation of a mathematical text from French to Russian, remarks:

"This report was especially interesting for those who knew and recalled the prehistory of machine translation, going back to 1939. In that year P. P. Smirnov-Troyanskiy, an inventor, appeared in the offices of the Academy of Sciences USSR and said that he had worked out a machine translation method; he asked that his invention be studied from the linguistic point of view. P. P. Smirnov-Troyanskiy's invention was met with profound skepticism by linguists; it was regarded as impractical and completely unnecessary. Only a few conceded that machine translation was possible. In the course of subsequent discussions, in which I took part, it turned out that P. P. Smirnov-Troyanskiy had not related his machine translation methods to the concept of electronic computers, but that his method did make it possible to translate, say, a Russian text in Moscow and to issue it in French translation, say, in Paris. And if it were possible to use machines for translation into the languages of the peoples of the Soviet Union, we could so to speak 'circularize' a given text simultaneously in several languages. The matter dragged on until a high-level conference of specialists in the fields of mechanics, electrical engineering, and linguistics met 31 July 1944 in the Institute of Automation and Telemechanics, Academy of Sciences USSR. Unfortunately the specialists in mechanics and engineering sought to offer proof of the 'impossibility' of machine translation and, infringing on the field of linguistics, spoke about synonyms, about fine shades of meaning, in short, about things that had no relation to their field. As a result an exper-

[1] Dostert, Leon. "The Georgetown-IBM Experiment," Machine Translation of Languages. New York: John Wiley & Sons and The Technology Press of the Massachusetts Institute of Technology, 1955, pp. 124-135,

Sheridan, Peter. "Research in Language Translation on the IBM Type 701," IBM Applied Science Division Technical Newsletter No. 9, January, 1955, pp. 5-24.

imental model of the translating machine (with a dictionary tabulator of 1000 words) was never built. P. P. Smirnov-Troyanskiy left Moscow soon afterward, so far as I know, and has since died, according to the information I have received."[2]

It is obvious from Zhirkov's representation of the Smirnov-Troyanskiy episode that he by no means intended to register by it a claim of Russian priority. In fact, Smirnov-Troyanskiy's ideas do not seem to have played any role in the thinking of MT pioneers in Russia or elsewhere. The earliest instance of suggestions and of the investigation of ideas of lasting importance and consequence for MT seems to have been the discussions that took place in 1946 between Dr. Warren Weaver of the Rockefeller Foundation and Dr. A. D. Booth of the London Birkbeck College Computation Laboratory. Weaver, thinking of different languages as some sort of different code systems, pointed to the techniques developed during World War II for the breaking of enemy codes and suggested the use of such techniques for the detection of basic features which might be common to all languages. Booth, on the other hand, thought that any digital computing machine could be used for MT if it had the necessary storage capacity and if a mere word-for-word or dictionary translation was considered sufficient.[3] In 1947, Booth and D. H. V. Britten collaborated at the Institute for Advanced Study at Princeton in the elaboration of details for such a dictionary translation by a digital computer. The results of their work are mentioned in a report to the Rockefeller Foundation at the end of that year.[4]

One of the prerequisites in Booth's scheme had been the availability in a digital computer of a permanent memory device with a capacity sufficiently large to permit the storage of a bilingual operational lexicon. In those early days of MT research memory devices with such a storage capacity did not yet exist. It was, however, pointed out that, if one limited MT to texts belonging to only a branch or even a subbranch of science, a much smaller bilingual vocabulary of perhaps manageable proportions may be involved. This idea of a "microglossary"[5] was later taken up by Dr. Victor A. Oswald of the University of California in Los Angeles[6] and subsequently followed up by others in their MT research.[7]

Dr. R. H. Richens of the Commonwealth Bureau of Plant Breeding and Genetics, Cambridge, England, thought of another way to solve the problem of limited permanent storage. Early in 1948 he suggested the compilation of MT dictionaries whose lexical units consisted entirely of word stems and of all possible endings, instead, as is usually the case, of the complete words.[8] He collaborated with Booth in testing these ideas extensively with sample dictionaries for ten languages used by a human ignorant of these languages who simulated the operations of a machine supplying word-for-word translations.[9]

The next phase in the history of MT was initiated by a memorandum, entitled Translation, which Dr. Warren Weaver wrote on July 15, 1949, and sent to some 200 scholars working in various fields. This memorandum strongly suggested the possibility of using electronic computers for the high-speed mass translation of scientific material, mentions his attempt to interest Professor Norbert Wiener of M.I.T., reports on the proposals of Booth and Richens, and, fully aware of the complexity of the linguistic problems involved, outlines several lines of attack in which due consideration is given to the phenomenon of universals, the problem of multiple meaning, and the "pinpointing" function of environment or context.[10]

There can be no doubt about the far-reaching significance of this memorandum for the rapid development and spread of MT research in the USA, Europe, USSR and Japan. Other important milestones in the short but already very promising history of the new field of linguistic research aimed at the automation of the translation process are (in this order):

1. The First International Conference on Machine Translation held on June 17-20, 1952, at M.I.T.

---

[2]Voprosy Yazykoznaniya (Problems of Linguistics), Vol. V, No. 5, September-October, 1956, Moscow, pp. 121-124. The translation is quoted from page 31 of the booklet marked JPRS/DC-68, CSO DC-1750, and circulated by The Office of Documentation of the National Science Foundation, Washington, D. C.

[3] William N. Locke and A. Donald Booth, Historical Introduction,"Machine Translation of Languages, p. 2. Cf. also Andrew D. Booth, L. Brandwood and J. P. Cleave, Mechanical Resolution of Linguistic Problems, Academy Press, Inc., New York, 1958, p. 2.

[4]Machine Translation of Languages, pp. 2 and 3.

[5]This term, created by Oswald for these small bilingual dictionaries, is still much used in MT literature. The present author suggested "idioglossary" as more appropriate because what is meant is the vocabulary peculiar to a branch or subbranch of knowledge, and its foreign correspondences.

[6]a) Victor A. Oswald, Jr., and Richard H. Lawson, "An Idioglossary for Mechanical Translation," Modern Language Forum, Vol. 38, No. 3-4, pp. 1-11, September-December, 1953. b) Victor A. Oswald, Jr., "The Rationale of the Idioglossary Technique," Research in Machine Translation, Report of the Eighth Annual Round Table Meeting on Linguistics and Language Study, Georgetown University Press, Washington, D. C., 1957.

[7]For example, the Georgetown University MT project.

[8]Machine Translation of Languages, p. 3.

[9]a) Mechanical Resolution of Linguistic Problems, p. 2, b) A. D. Booth and R. H. Richens, "Some Methods of Machine Translation," Machine Translation of Languages, pp. 24-46.

[10] Warren Weaver, "Translation," Machine Translation of Languages, pp. 15-23.

[11]a) Erwin Reifler,"Report on the First Conference on Mechanical Translation," MT, Vol. 1, No. 2, August, 1954, pp. 23-32. b) A. C. Reynolds, Jr., "The Conference on Mechanical Translation," MT, Vol. 1, No. 3, December, 1954, pp. 47-55. c) Machine Translation of Languages, pp. 5 and 6.

2. A number of grants by the Rockefeller Foundation.[12]

3. The IBM-Georgetown University Experiment in January, 1954, mentioned earlier.[13]

4. The creation of the journal for MT development, called Mechanical Translation or MT, in 1954, edited by Dr. Victor H. Yngve and published at M.I.T.[14]

5. The design and construction of the first pilot model of a German-English translation machine in 1954 at the University of Washington in Seattle, Washington, USA. In view of subsequent developments, its value is now only that of a candidate for a future MT museum. It has, however, served an important purpose in focusing at the University of Washington the interdepartmental endeavours in the new field of Linguistic Engineering or Engineering Linguistics.[15]   The pilot model was planned by Dr. T. M. Stout and built by graduate students of the Electrical Engineering Department of the University.[16]

6. The publication of the first book on MT, a collection of essays by 14 pioneers, in 1955 under the title Machine Translation of Languages.[17]

7. The initiation in 1955 of theoretical and experimental work in MT in the Soviet Union.[18]

8. The national meeting of MT pioneers with representatives of U. S. government agencies on January 18, 1956, in Cleveland, Ohio, during the Conference on the Practical Utilization of Knowledge at Western Reserve University.

9. The Second International Conference on Machine Translation at M.I.T. on October 20, 1956, attended by some thirty workers in the field from the United States, Canada, and England. For the first time papers on MT were also submitted by researchers in the Soviet Union.[19]

10. The development of a photoscopic memory device with a practically unlimited permanent storage capacity and the shortest access time attained so far,[20]

11. The initiation and distribution since July, 1957 of the publication Current Research and Development in Scientific Documentation compiled by the Office of Science Information Service of the National Science Foundation in Washington 5, D. C. It is now published semi-annually and lists all pertinent U. S. and foreign projects coming to the attention of the staff of the Foundation, including all MT projects.[21]

12. The financing by U. S. government agencies of MT projects in academic institutions and industrial organizations for the solution of the linguistic and engineering problems involved,[22] and the employment of modern structural linguists for this purpose by a number of industrial organizations.

13. The academic recognition accorded MT research in the form of special sessions on MT included in the programs of scientific societies.[23]

---

[12]The present author has been the grateful recipient of two Rockefeller grants in 1952 and 1953.

[13]Cf. page 1 of this "Historical Outline," and footnote No. 1.

[14]Address for editorial correspondence: Dr. Victor H. Yngve, Room 20B-101, Department of Modern Languages and Research Laboratory of Electronics, M.I.T., Cambridge 39, Massachusetts, U.S.A.

[15]Reifler, Erwin. Materials for the MT Pilot Model at the University of Washington. (mimeographed report).

[16]a)Stout, T. M. "Computing Machines for Language Translation," Mechanical Translation, Vol. 1, No. 3, December, 1954, pp. 41-46.

  b)Douthwaite, Geoff. "The University of Washington Automatic Language Translator," Washington Engineer, Vol. VIII, No. 4, February, 1956.

[17]Cf. footnote No. 1.

[18] a)Rozentsveig, V. Yu. "The Work in the Soviet Union on Machine Translation from Foreign Languages into Russian and from Russian into Foreign Languages," IV International Congress of Slavicists, Reports, Moscow, 1958. Translated by Lew R. Micklesen, mimeographed copy, p. 1. (Obtainable from the National Science Foundation.)

  b)Kulagina, O. S., A. I. Martynova, and T. M. Nikolayeva in Mechanical Translation at the Academy of Sciences of the USSR, V. A. Steklov Institute of Mathematics of the Academy of Sciences of the USSR, Moscow, 1958, page 3, claim that 1954 was the year in which MT research was started in the Soviet Union by the Academy of Sciences.

[19]See MT, Vol. III, No. 2, November, 1956, for the reports of the different research groups.

[20]King, Gilbert W. "The Photoscopic Memory System," Final Report on Computer Set AN/CSO-16(XW-1), Vol. 1. IBM Research Center, Yorktown Heights, New York, June 20, 1959.

[21]The reports appear in April and October. The last issue at the time this chapter was being prepared is dated October, 1959. For sale by the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Price 15 cents. Scientists and other individuals concerned with scientific documentation are placed on the mailing list for these reports on request.

[22]Cf. Current Research and Development in Scientific Documentation for details about these MT projects.

[23]For example, at the annual meetings of the American Oriental Society at Princeton University, April, 1957, at the VIII International Congress of Linguists at Oslo University, Oslo, Norway, in August, 1957 (See the Proceedings of the Congress, Oslo University Press, Oslo, 1958, pp. 502-539), and at the IV International Conference of Slavicists in Moscow, USSR, in September, 1958 (cf. the forthcoming volume of abstracts of papers presented at the Conference).

14. The First All-Union Conference on Machine Translation, Moscow, USSR, May 15-21, 1958.[24]
15. The International Conference on Information Processing and Auto-Math under the sponsorship of UNESCO, Paris, June 13-23, 1959, whose program included a session on Automatic Translation of Languages.
16. The International Conference for Standards on a Common Language for Machine Searching and Translation, Cleveland, Ohio, September 6-12, 1959.[25]

According to the latest (October, 1959) issue of Current Research and Development in Scientific Documentation, MT research in the United States is at present being pursued at the following academic institutions (in alphabetic order):

1. University of California, Computer Center, Berkeley, California.[26]
2. Georgetown University, Institute of Languages and Linguistics, Washington, D.C.[27]
3. Harvard University, Computation Laboratory, Cambridge, Massachusetts.[28]
4. Massachusetts Institute of Technology, Research Laboratory of Electronics, Cambridge, Massachusetts.[29]
5. University of Michigan, Willow Run Laboratories, Ann Arbor, Michigan.[30]
6. University of Washington, Department of Far Eastern and Slavic Languages and Literature and the Electrical Engineering Department, Seattle, Washington.[31]
7. Wayne State University, Department of Slavic Languages and Computer Center, Detroit, Michigan.[32]

MT research is at present also being carried on in the following United States organizations:
1. The National Bureau of Standards, Applied Mathematics Division, Washington, D. C.[33]
2. Ramo-Wooldridge, Inc., Machine Translation Department, Los Angeles, California.[34]
3. Rand Corporation, Santa Monica, California.[35]

Outside the United States, MT research teams are at present working in England, Israel, Italy, Japan, and the Soviet Union, namely:

A. <u>England</u>

1. Birkbeck College, Department of Numerical Automation, London.[36]

---

[24]Cf. <u>Abstracts of the Conference on Machine Translation, Moscow,</u> May 15-21, 1958, 71 papers. Photocopies of the English translation of this report, labeled JPRS/DC-241, may be purchased from the Photoduplication Service, Library of Congress, Washington 25, D. C.

[25]Information about the numerous papers on MT and the Proceedings, scheduled to be published in March, 1960, by Interscience Publishers, Inc., New York, may be obtained from Dr. Allen Kent, Associate Director, Center for Documentation and Communication Research, School of Library Science, Western Reserve University, Cleveland 6, Ohio.

[26]Russian-English. Principal Investigator: Sidney M. Lamb. Supported by The National Science Foundation.

[27]Russian-English, English-Arabic, English-Chinese; preliminary work in the development of a general trans-Slavic syntax starting with Czech, Polish and Serbo-Croatian. Director: L. E. Dostert. (For publications see Bibliography). Supported by The National Science Foundation.

[28]Russian-English. Principal Investigator: Anthony G. Oettinger. (For publications see Bibliography). Supported by The National Science Foundation.

[29]German-English. Principal Investigator: Victor H. Yngve. (For publications see Bibliography). Supported by The National Science Foundation.

[30]Russian-English. Principal Investigator: Andreas Koutsoudas. (For publications see Bibliography). Supported by The Army Signal Corps.

[31]German-English, Russian-English. In the planning stage: Chinese-English. Project Director: Erwin Reifler. (For publications see Bibliography). Supported by The Air Force.

[32]Russian-English. Principal Investigators: H. H. Josselson and A. W. Jacobson. Supported by The Office of Naval Research.

[33]Russian-English. Principal Investigator: I. Rhodes. Supported by Office of Ordnance Research, Department of the Army.

[34]Russian-English. Principal Investigator: Don. R. Swanson. (For publications see Bibliography). Supported by The Air Force.

[35]Russian-English. Principal Investigator: David G. Hays. (For publications see Bibliography). Supported by The Air Force.

[36]French-English, German-English. Principal Investigator: Andrew D. Booth. Supported by The Nuffield Foundation.

2. Cambridge Language Research Unit, 20 Millington Road, Cambridge.[37]

3. National Physical Laboratory, Control Mechanisms and Electronics Division, Teddington, Middlesex.[38]

B. Israel

1. Hebrew University, Jerusalem.[39]

C. Italy

1. Università Degli Studi Di Milano, Centre Di Cibernetica E Di Attività Linguistiche.[40]

D. Japan

1. Kyoto University, Department of Psychology, Faculty of Letters.[41]
2. Electrotechnical Laboratory, Electronics Division, Tokyo.[42]

E. Soviet Union[43]

1. Academy of Sciences USSR:
   a. Institute of Precision Mechanics and Computer Technology (ITM and VT).[44]
   b. Steklov Mathematical Institute (MIAN).[45]
   c. Laboratory of Electrical Modelling of the All-Union Institute of Scientific and Technical Information (LE).[46]
2. First Moscow State Institute of Foreign Languages (I. MGPIIYA).[47]
3. Leningrad State University, Experimental Laboratory of Machine Translation (ELMP).[48]
4. Gorky State University.[49]
5. Academy of Sciences of the Armenian SSR, Computation Center.[50]
6. Academy of Sciences of the Georgian SSR, Institute of Automation and Telemechanics.[51]

---

[37]General MT programs for many-to-one translations by means of an automatic thesaurus. Italian-Interlingual and Interlingual-English dictionary. Principal Investigator: Margaret Masterman. (For publications see Bibliography). Supported by The U. S. National Science Foundation.

[38]Russian-English. Project Director: D. W. Davies.

[39]Theoretical studies into the structure of language, linguistic models, etc. Principal Investigator: Yehoshua Bar-Hillel. (For publications see Bibliography).

[40]Russian-English, Italian-English. The approach is based on the research results of the Italian operational school concerning the thought processes underlying language and translation. Principal Investigator: Silvio Ceccato. (For publications see Bibliography). Supported by The United States Air Force.

[41]English-Japanese. Principal Investigator: Ichiro Honda.

[42]English-Japanese. Principal Investigator: H. Wada. (For publications see Bibliography).

[43]The information concerning MT in the Soviet Union is compiled from V. Yu. Rozentsveig, op. cit. (footnote 18) and O. S. Kulagina, A. I. Martynova and T. M. Nikolayeva, op. cit. (Ibidem; for the names of the many investigators we have to refer the reader to these two publications. For other publications see Bibliography).

[44]From English, German, Chinese, and Japanese into Russian.

[45]From English, French and Hungarian into Russian. Director of Research: A. A. Lyapunov.

[46]Isolation and cataloguing of the system of relationships (syntagmas) in the Russian language for MT purposes.

[47]From Russian into English, French and Spanish. Director of Research: I. I. Revzin.

[48]From Indonesian, Arabic, Hindi, Japanese, Burmese, Norwegian, English, Spanish and Turkish (this is the order given by Rozentsveig) into Russian. Director of Research: N. D. Andreyev. In the report to the International Conference for Standards on a Common Language for Machine Searching and Translation in Cleveland, Ohio, September 6-12, 1959, entitled "Basic Directions of Work at the Experimental Laboratory of Machine Translation," Andreyev mentions the following languages for which algorithms of independent analysis are being developed by his laboratory (the order is his): Russian, Chinese, Czech, German, Rumanian, Vietnamese, Serbo-Croatian, English, French, Spanish, Norwegian, Arabic, Hindustani, Japanese, Indonesian, Burmese, and Turkish.

[49]Both Russo-Foreign and Foreign-Russian MT.

[50]Armenian-Russian and Russian-Armenian.

[51]Georgian-Russian and Russian-Georgian.