

THE ORIGINS OF MT

Andrew D. Booth and Kathleen H.V. Booth
Autonetics Research Associates
Sooke, BC V0S 1N0 Canada
aboath@ios.bc.ca
Tel: (250) 642 5352

1. Origins

In his 1666 dissertation, *The Art of Combinations*, Leibniz had clearly indicated in his discussion of the possible mechanization of both arithmetic and thought processes, that his logical processor could be used to transform one language into another. He also considered the extraction of ideas from text and their expression in terms of a metalanguage (Saw 1954). The culmination occurred in his design of the first practicable mechanical multiplying machine (1694) and in a binary multiplier (Eriksson et al. 1996).

Another early claim, by the Russians, was to the effect that a Soviet engineer had produced a translating machine in the 1930s. This appears unlikely because MT depend heavily on a large and complex dictionary—something which was not even on the horizon at that time.

We date realistic possibilities starting with two meetings held in 1946. The first was between Warren Weaver, Director of the Natural Sciences Division of the Rockefeller Foundation, and Norbert Wiener. The second was between Weaver and A.D. Booth in that same year. The Weaver-Wiener discussion centered on the extensive code-breaking activities carried out during the World War II. The argument ran as follows: decryption is simply the conversion of one set of "words"—the code—into a second set, the message. The discussion between Weaver and A.D. Booth on June 20, 1946, in New York identified the fact that the code-breaking process in no way resembled language translation because it was known a priori that the decrypting process must result in a unique output. The main purpose of this meeting, however, was to interest the Rockefeller Foundation in supporting development of an electronic computer at the University of London.

Things remained thus until the end of a visit to the Institute for Advanced Study, Princeton, in 1947. On May 20 of that year another talk with Weaver showed that the Foundation was unlikely to fund a machine at London for numerical work but that a submission indicating interest in MT would be well received. This was duly submitted and was successful. The core of the proposal involved the use of the high-speed storage organ as an electronic dictionary.

2. London, 1948-1962

On our return to London, work was at once started on two fronts: ADB designing and supervising the construction of the new computer and KHVB directing the development of suitable programs. It must be noted that in those days all coding was done in machine language (not even "assembler") and had to be inserted via switches

Some code for a computerized dictionary was tested and found to be very effective. The lookup method used from the first was binary partitioning—a technique which we regarded as so obvious that we did not deem it worth publication until we discovered that several major computer groups were using sequential search strategies (A.D. Booth 1955). More extensive descriptions of our work are given in *Automatic Digital Calculators* (A.D. Booth & K.H.V. Booth 1953) and in *Programming for an Automatic Digital Computer* (K.H.V. Booth 1958). The latter contains a complete MT program of that era.

Other activities of our group included work on the "translation" of English to Braille by J.P. Cleave (1956), language processing in general by M. Levison (1962), and the chronology of the Plato dialogues by L. Brandwood. Much of the work is described in *Mechanical Resolution of Linguistic Problems* by A.D. Booth, L. Brandwood, and J.P. Cleave (1958). Our thanks are due the Nuffield Foundation, which provided major funding during this period.

3. Saskatchewan, 1962-1972

In 1962 we migrated to Canada and in due course became involved with the Queen's Printer in a most interesting project. One problem in Canada is the presence of two "official" languages: English and French. The parliamentary system requires that the Hansard, the daily transcript of the day's debates, be available in both languages the next day. This was the problem which we were asked to address. Our own view was, and still is, that output of good quality can be produced only with a post-editor. With the aid of linguists Corinne Brown and Gloria Geitz and programmer Charles Stock (K.H.V. Booth & Stock

1971) we produced such a program and demonstrated it at the National Research Council Laboratories in Ottawa in 1971. It is interesting to note that the only languages available on our University mainframe were FORTRAN and COBOL. The program was written in the latter.

4. Lakehead University, 1972-1978

When we migrated to Ontario in 1972, our work on MT was continued with emphasis on the construction of a large dictionary and the use of statistical techniques in linguistic analysis (A.D. Booth et al.).

5. Retirement, 1978-

Our work on MT terminated with our retirement in 1978 except for one small but interesting experiment. In 1979 the first personal computers became available and we were involved in a project to develop an energy management package. The machine chosen was a Commodore PET 2000 and, having some time to spare, ADB decided to write an MT program for this machine. Here the language was BASIC and a French-to-English program was soon produced and demonstrated to the Canadian National Research Council in 1980. The program incorporated most of the methods devised in our earlier work but, because of storage limitations, it had only a small dictionary. The output quality was equal to or better than that generated by currently available inexpensive "translation" programs. The main superiority of our program lay in the use of our category count technique (K.H.V. Booth 1970) for the resolution of such ambiguities as Chomsky's "The pen is in the box."

REFERENCES

- Booth, A.D. 1955. On the Use of a Computing Machine as a Mechanical Dictionary. *Nature*, 176:565.
- Booth, A.D., and Booth, K.H.V. 1953. *Automatic Digital Calculators*. 1st ed. London: Butterworths.
- Booth, A.D., Brandwood, L., and Cleave, J.P. 1958. *Mechanical Resolution of Linguistic Problems*. London: Butterworths.
- Booth, K.H.V. 1958. *Programming for an Automatic Digital Calculator*. London: Butterworths.
- Booth, K.H.V. 1960. Statistical Parsing by Computer. *Pensiero e Linguaggio* 1:107.
- Booth, K.H.V., and Stock, C. 1971. English-French Translation on a Computer. *Pensiero e Linguaggio*, 2:231.
- Cleave, J.P. 1956. The Application of an Automatic Digital Computing Machine to Problems of Data Transformation and Language. Ph.D. thesis, University of London.
- Eriksson, K., Estep, D., Hansbo, P., and Johnson, C. 1996. *Computational Differential Equations*. Cambridge: Cambridge University Press. p. 20.
- Levison, M. 1962. The Application of a Computing Machine to Linguistic Problems. Ph.D. thesis, University of London.
- Saw, Ruth Lydia. 1954. *Leibniz*. London: Penguin Books. p. 214.