[from: *Language Today*, no.16, January 1999, pp. 19-20]

Milestones in machine translation
John Hutchins

No.5: The IBM-Georgetown demonstration, January 1954

On the 8th January 1954, newspapers in the United States carried front-page reports of what may well have been surprising news for the great majority of its readers. They were reports of the first public demonstration of a MT system on the previous day. It was the result of a joint project by IBM staff under Cuthbert Hurd and members of the Institute of Linguistics at Georgetown University under Leon Dostert. It was the pilot demonstration of the practical feasibility of MT that Dostert had proposed at the first conference on machine translation two years earlier [see Milestones no.4].

The report in the *New York Herald Tribune* marvelled at the achievement:

> "A huge electric 'brain' with a 250-word vocabulary translated mouth-filling Russian sentences yesterday into simple English in less than ten seconds. As lights flashed and motors whirred inside the 'brain' the instrument's automatic type-writer swiftly translated statements on politics, law, science and military affairs. Once the Russian words were fed to the machine no human mind intervened."

For the *New York Times* it represented "the cumulation of centuries of search by scholars for 'a mechanical translator'." The reporter laid particular stress on the fact that it had been accomplished on "a standard commercial model of the largest International Business Machines 'stock' computer" – in fact, the IBM 701, the first electronic computer to become widely available, and which had been put on the market only the previous April. The reporter gave a detailed account of how Russian sentences were punched onto IBM 80-column cards by an operator, put into the machine, and English sentences came out. The linguistic processes were scarcely alluded to. However, the limitations of this trial program were readily conceded. Nevertheless the demonstrators were prepared to predict that "such a device should be ready within three to five years, when the Georgetown scholars believe they can complete the 'literary' end of the system," and optimistic claims were made that other languages would be easily added. "As soon as cards for Russian are completed, sets will be made for German and French. Then other Slavic, Germanic and Romance languages can be set up at will."

The next day, the *Christian Science Monitor* reflected some of the popular expectations of computers at the time. Much was made of the speed and 'intelligence' of the computer:

> "The girl who operated the 701 did not understand a word of Soviet speech and yet more than 60 Soviet sentences were given to the 'brain' which translated smoothly at the rate of about 21/2 lines a second… The 'brain' didn't even strain its superlative versatility and flicked out its interpretation with a nonchalent attitude of assumed intellectual achievement."

Again, Dostert's predictions were also reported: "Although he emphasised it is not yet possible 'to insert a Russian book at one end and come out with an English book at the other', the professor forecast that 'five, perhaps three, years hence, interlingual meaning conversion by electronic process in important functional areas of several languages may well be an accomplished fact'."

Two days later the *Christian Science Monitor* carried an editorial (one of the few ever devoted to MT), which provided a more sober assessment:

> "Such an accomplishment, of course, is far from encompassing the several hundred thousand words which constitute a language. And with all the preparations for coping with syntax, one wonders if the results will not sometimes suggest the stiffness of the starch mentioned in one of the sentences as being produced by mechanical methods. Nevertheless, anything which gives promise of melting some of the difficulty which writers and speakers of different languages encounter in understanding each other – particularly as between English and Russian today – is certainly welcome."

The demonstration had attracted a great deal of attention. The general public was led to expect automatic translation of unrestricted texts within the next few years – expectations, which were of course to be continually disappointed. More significantly, however, US government agencies were encouraged to support research on a large scale for the next decade, and MT groups were established in universities throughout the United States. Similarly, the Soviet authorities had been alerted to the possibilities, and research groups began to appear in many locations in the Soviet Union.

However, the enthusiasm was not well founded. On the technical side, there were still major impediments. Peter Sheridan, the IBM staff member responsible for the programming, has given a fascinating account of what was involved (*IBM Technical Newsletter* 9, January 1955, pp.5-24). As one of the first non-numerical applications of computers, every aspect of the process was unknown territory. Decisions had to be made about how to code alphabetic characters, how to transliterate Russian letters, how the Russian vocabulary was to be stored on a magnetic drum, how the 'syntactic' codes were to operate, how much information was to go on each punched card, etc. Detailed flowcharts were drawn up for what today would be simple in-built operations of a computer program. Some years were to pass before researchers could program with 'assembly language' codes, and many more before they had available 'high-level' programming languages.

On the linguistic side, the Georgetown researchers had themselves emphasised the limited nature of the project. Paul Garvin gave a full account somewhat later ('The Georgetown-IBM experiment of 1954: an evaluation in retrospect' *Papers in linguistics in honor of Léon Dostert* (The Hague: Mouton, 1967), pp.46-56). They had carefully selected a corpus of 49 Russian sentences, with a lexicon of just 250 words. The lexical items were coded in a system of 'digital diacritics' of three types. One series (Program Initiating Diacritics) indicated which of the six operational rules was to be applied. A second (Choice Determining Diacritics) indicated what contextual information should be sought to determine selection of output. And the third series (Address Diacritics) indicated the storage locations of English equivalents.

The six operational rules were (using the original numbering):

0. No problems of selection: there is one-to-one equivalence of source and target words, and the word order of the source is to be followed.
1. There is a change of order: the words are to be inverted.
2. The choice between target equivalents is determined by an indication ('diacritic') in the following word.
3. The choice of target word is determined by an indication in the preceding word.
4. The word in the source is omitted, and no word appears in the target sentence.
5. A word is inserted in the target which has no correspondent in the source.

Despite its limitations, Garvin considered the experiment "realistic" because "the rules dealt with genuine decision problems, based on the identification of the two fundamental types of translation decisions: selection decisions and arrangement

decisions." What were the limitations? Firstly, there were three major ones: a restriction of the search span to immediately adjacent items, the restriction of target words to just two possibilities, and the rearrangements to two immediately adjacent items. Secondly, there were severe constraints on the translation of Russian case endings: either a case suffix was not translated at all or it was translated by one "arbitrarily assigned" English preposition. Thirdly, there were further limitations inherent in the selection of the sentences themselves. All sentences were declaratives; there were no negatives, no interrogatives, no coordinate or subordinate clauses, and all verbs were in the third person. Finally, the English articles were inserted *ad hoc* to fit the particular words of the corpus.

Such limitations made it possible for the output to be impressively and deceptively idiomatic. Other MT researchers at the time were highly critical of the Georgetown team, believing that it was at best premature to demonstrate any MT system, and certainly that it was misleading to produce such high quality output. It was inevitable that the public and the funding agencies would expect future MT output to be of the same quality, indeed that they would expect it to achieve near-human standards within a few years. Such expectations could not possibly be met. When ten years later, MT systems seemed to be no better than in this 1954 demonstration, the US agencies set up the Automatic Language Processing Advisory Committee (ALPAC) which in its 1966 report was to conclude that support for MT should not continue in the United States.

In retrospect, this was one of the most unfortunate consequences of the IBM-Georgetown demonstration of 1954. Against this, however, it has to be admitted that it was the first actual implementation of an automatic translation system of some kind. All previous work on MT had been theoretical in the sense that none of the proposals had been implemented as 'real' computer programs. Some had been simulated on punched cards or paper slips, others were no more than 'thought experiments'. In view of the still primitive stage of computing, it was indeed remarkable that anything resembling automatic translation could be achieved at all at this date. Despite all its limitations and the controversy surrounding it, the 1954 demonstration marked the beginning of MT as a reality. Just seven years after Weaver had first proposed the use of computers to translate, here was some evidence for its feasibility.