Milestones in Machine Translation

Part 2 - Warren Weaver's 1949 memorandum

W. John Hutchins continues his history of MT

In July 1949, Warren Weaver sent to some 30 acquaintances a memorandum on the possibilities of using the newly invented digital computers for the task of translating documents. He had first mentioned the possibility (as recounted in the first article of this series, *LT3*) in March 1947 in a letter to Norbert Wiener and in conversation with Andrew Booth, a British scientist. In the next two years, he was urged by his colleagues at the Rockefeller Foundation to elaborate his ideas. The result was a memorandum, entitled simply 'Translation'.*

Warren Weaver was director of the Natural Sciences Division of the Rockefeller Foundation, where he was responsible for instigating and approving grants for major projects in molecular engineering and genetics, in agriculture - particularly for developing new strains of wheat and rice in Central and South America and in South East Asia - and in medical research. He was a mathematician with special interest in probability and statistics. During the war, Weaver had been seconded from the Foundation to be the chair of the Applied Mathematics Panel of the US Office of Scientific Research and Development, and was responsible for directing the work of several hundred mathematicians on 'operations research' of all kinds. As such, he was fully familiar with the development of electronic calculating machines, and well aware of the successful application of mathematical and statistical techniques in the decipherment of enemy messages. The impact of his memorandum, which effectively launched research in machine translation in the United States, was attributable to Weaver's influence on the major policy makers in US government agencies, and his widely recognised expertise in mathematics and computing.

After stating "the obvious fact that a multiplicity of languages impedes cultural interchange between the peoples of the earth, and is a serious deterrent to international understanding", Weaver described the origins of his own interest in the topic. He had been impressed at the success of cryptography based on, as he put it, "frequencies of letters, letter combinations, intervals between letters and letter combinations, letter patterns, etc. which are to some significant degree independent of the language used" (this is Weaver's own underlining in the typescript). He had noticed also a paper by a Sinologist, Erwin Reifler, who had remarked that "the Chinese words for 'to shoot' and 'to dismiss' show a remarkable phonological and graphic agreement", to which Weaver added: "This all seems very strange until one thinks of the two meanings of 'to fire' in English. Is this only happenstance? How widespread are such correlations?"

He continued with a brief account of what had been done already. Firstly, there were some experiments with punched cards by Richard Richens and Andrew Booth in England, which had produced crude word-for-word translations of scientific abstracts. Secondly, there had been newspaper reports of a computer in Los Angeles which was intended to be used for simple experiments in translation (although Weaver does not say so, the computer was based at the Institute for Numerical Analysis at the University of California at Los Angeles (UCLA), a branch of the US National Bureau of Standards, and the research was directed by Harry Huskey who had previously worked on computers at Princeton University and at the National Physical Laboratory in England). These were, of course, just the beginnings, and Weaver was quick to point out the grave limitations of any simplistic word for word approach. His memorandum was designed to suggest more fruitful methods.

He put forward four proposals. The first was that the problem of multiple meanings might be tackled by examinations of immediate contexts.

"If one examines the words in a book, one at a time through an opaque mask with a hole in it one word wide, then it is obviously impossible to determine, one at a time, the meaning of words. 'Fast' may mean 'rapid'; or it may mean 'motionless'; and there is no way of telling which.

"But, if one lengthens the slit in the opaque mask, until one can see not only the central word in question but also say N words on either side, then, if

www.logos.it/language_today

N is large enough one can unambiguously decide the meaning..."

The problem was, of course, to determine how much context would be required - and this might vary from subject to subject. However, Weaver thought that "relatively few nouns, verbs and adjectives" were actually ambiguous, so that the problem was not large. How wrong he was!

His second proposal started from the assumption that there are logical elements in language. It drew attention to a theorem proved by McCulloch and Pitts - developed in fact in the context of research on the neural structure of the human brain - that "a robot (or computer) constructed with regenerative loops of a certain formal character is capable of deducing any legitimate conclusion from a finite set of premises". The mathematical possibility of computing logical proofs suggested to Weaver that "insofar as written language is an expression of logical character" the problem of translation is formally solvable.

The third proposal took up again the applicability of cryptographic methods. It involved the recent 'information theory' of Claude Shannon (Weaver was writing a book about it with Shannon at the time). The theory is concerned with the basic statistical properties of communication, including the effects of noise in telecommunication channels and of relative frequencies of signals. In particular, it embraced "the whole field of cryptography" (Shannon was himself the author of one of the most influential reports on the topic, which had remained classified until 1949). Weaver admitted that the validity of the cryptographic approach was difficult to assess, but he was obviously attracted:

"It is very tempting to say that a book written in Chinese is simply a book written in English which was coded into the 'Chinese code'. If we have useful methods for solving almost any cryptographic problem, may it not be that with proper interpretation we already have useful methods for translation?"

As it happens, it was not long before researchers in machine translation recognised the fallacy of the argument. The mistake lay in the confusion between the activities of decipherment and translation, which arise whenever the same person does both - as indeed is often the case.

For his fourth proposal, Weaver became more utopian. It was based on the belief that, just as there may be logical features common in all language, there may also be linguistic universals. He ended his memorandum, therefore, with one of the best known metaphors in the literature of machine translation:

"Think, by analogy, of individuals living in a series of tall closed towers, all erected over a common foundation. When they try to communicate with one another, they shout back and forth, each from his own closed tower. It is difficult to make the sound penetrate even the nearest towers, and communication proceeds very poorly indeed. But, when an individual goes down his tower, he finds himself in a great open basement, common to all the towers. Here he establishes easy and useful communication with the persons who have also descended from their towers.

"Thus it may be true that the way to translate from Chinese to Arabic, or from Russian to Portuguese, is not to attempt the direct route, shouting from tower to tower. Perhaps the way is to descend, from each language, down to the common base of human communication - the real but as yet undiscovered universal language - and then re-emerge by whatever particular route is convenient."

Weaver realised, of course, that this approach involved a "tremendous amount of work in the logical structures of languages before one would be ready for any mechanization." However, he believed that some steps towards it had been made, particularly in the proposed Basic English of Ogden and Richards, which was then at the height of its popularity.

The reception of the memorandum was mixed. Some rejected the very idea of mechanising the complexity of translation - in much the same terms as many professional translators reject machine translation today - but others were prepared to give some thought to the possibility. One of the first was Erwin Reifler, the Sinologist referred to by Weaver. Within the next few months, Reifler wrote studies of how crude word-for-word renditions could be made use of, introducing ideas on 'pre-editing' and 'post-editing' and the use of regularised languages. Another was Abraham Kaplan at the Rand Corporation, who followed up Weaver's suggested statistical approach to resolving problems of multiple meaning. There were also the beginnings of investigations of syntactic analysis at UCLA, but perhaps the most significant outcome was the decision at the Massachusetts Institute of Technology to appoint Yehoshua Bar-Hillel to a research position. Its impact will be the topic of the next article in this series. ■

* *Reproduced in: Locke, W.N. and Booth, A.D. (eds.) Machine translation of languages: 14 essays (Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology, 1955).*