APPENDIX IV

A DEMONSTRATION OF THE NON-FEASIBILITY OF FULLY-AUTOMATIC HIGH-QUALITY MACHINE TRANSLATION*

Y. BAR-HILLEL
Hebrew University, Jerusalem, Israel

One of the reasons why we do not as yet have any translation centers, not even *in* the planning stage, in which electronic computers, general or special purpose, are used to automate certain parts of the translation process, in spite of the fact that such centers would fulfill a vital function in saving a considerable amount of qualified human translator time per document translated, and thereby facilitate more, quicker and, after some time, cheaper translation, is the reluctance of many MT workers to recognize that their pet idea of inventing a method for fully-automatic high-quality machine translation (FAHQMT) is just a dream which will not come true in the foreseeable future. By not realizing the practical futility of this aim, whatever its motivational importance for certain types of basic research, they have managed to fool themselves and the agencies which sponsored their research not to be satisfied with a partly automated translation system whose principles are well understood today, but to wait for the real thing which was believed, and made believe, to be just around the corner.

During the last year, I have repeatedly tried, through personal talks, lectures before conferences and small groups, as well as in articles[1)], to point out the illusoriness of the FAHQMT ideal already in respect to mechanical determination of the syntactical structure of a given source language sentence. These efforts of mine were based on certain deep theoretical insights into linguistic structure recently obtained by Chomsky[2)]. Today, I shall show that there exist extremely simple sentences in English — and the same holds, I am sure, for any other natural language — which, within certain linguistic contexts would be uniquely and unambiguously translated into, say, French or German or Russian or what have you by anyone with a sufficient knowledge of the two languages involved, though I know of no program that would enable a machine to come up with this unique rendering unless done so by a completely arbitrary and ad hoc procedure whose futility would show itself in the next example. *I* defy any of the MT experts gathered here — and anybody else — to show me where I am wrong. In case they are unable to do this, let me suggest that they stop talking about the attainability of FAHQMT, and thereby give the green light to those agencies and people who are interested in overcoming the severe problems created by the shortage of qualified human translators. If they themselves are not willing to cooperate in the establishment of a working, partly-automated, high-quality translation outfit because of the small amount of intellectual satisfaction that will accompany

---

*) This paper was prepared under a more general research program supported by the Office of Naval Research, Information Systems Branch, Contract No. Nonr-2578(00). NR 049-130, It is to be read before the International Conference for Information Processing, Paris, June 1959.

such an achievement, I for one understand this attitude fully and even share it to such a degree that I have no intention to spend my time on such a project. Let them, and myself, by all means go on and investigate the countless number of basic problems with a real challenge in them, concerning language models, special-purpose program languages for translation, machines that will learn to translate, etc. But I do not think they should discourage other people eager to establish a system, any system, that does save valuable man-power and does solve an urgent problem, by threatening them, so to speak, with the outlook of having to face tomorrow an incomparably better system that will make their efforts and achievements look childish and pointless.

I now come to my sentence. It is just:

The box was in the pen.

The linguistic context from which this sentence is taken is, say, the following:

Little John was looking for his toy box. Finally he found it.
The box was in the pen. John was very happy.

Assume, for simplicity's sake, that 'pen' in English has only the following two meanings: (1) a certain writing utensil, (2) an enclosure where small children can play. I now claim that no existing or imaginable program will enable an electronic computer to determine that the word 'pen' in the given sentence, within the given context has the second of the above meanings, whereas every reader with a sufficient knowledge of English will do this "automatically". Incidentally, we realize that the issue is not one that concerns translation proper, i.e., the transition from one language to another, but a preliminary stage of this process, i.e., the determination of the specific meaning in context of a word which, in isolation, is semantically ambiguous (relative to a given target-language, if you so wish).

It is an old prejudice, but nevertheless a prejudice, that taking into consideration a sufficiently large linguistic environment as such will suffice to reduce the semantical ambiguity of a given word. Let me quote from the memorandum which Warren Weaver sent on July 15, 1949 to some two hundred of his acquaintances and which became one of the prime movers of MT research in general and directly initiated the well-known researches of Reifler and Kaplan[3]: "...if...one can see not only the central word in question, but also say N words on either side, then, if N is large enough one can unambiguously decide the meaning of the central word. The formal truth of this statement becomes clear when one mentions that the middle word of a whole article or a whole book is unambiguous if one has read the whole article or book, providing of course that the article or book is sufficiently well written to communicate at all." Weaver then goes on to pose the practical question: "What minimum value of N will, at least in a tolerable fraction of cases, lead to the correct choice of meaning for the central word", a question which was, we recall, so successfully answered by Kaplan. But Weaver's seemingly lucid argument is riddled with a fateful fallacy: the argument is doubtless valid (fortified, as it is, by the escape-clause beginning with 'providing') but only for intelligent readers, for whom the article or book was written to begin with. Weaver himself thought at that time that the argument is valid also for an electronic computer, though he did not say so explicitly in the quoted passage, and on the contrary, used the word 'one'; that this is so, will be clear to anyone who reads with care the whole section headed "Meaning and Context". In this fallacious transfer

Weaver has been followed by almost every author on MT problems, including the Russian ones. It would be very easy to provide as many quotations as you wish to corroborate this statement of mine. But this is probably unnecessary since I do not believe that someone would wish to challenge me on this point.

Now, what exactly is going on here? Why is it that a machine with a memory capacity sufficient to deal with a whole paragraph at a time, and a sytactico-semantic program that goes, if necessary, beyond the boundaries of single sentences up to a whole paragraph (and, for the sake of the argument, up to a whole book, if you so wish) — something which has so far not gotten beyond the barest and vaguest outlines — is still powerless to determine the meaning of 'pen' in our sample sentence within the given paragraph? The explanation is extremely simple, and it is nothing short of amazing that, to my knowledge, this point has never been made before, in the context of MT, though it must surely have been made many times in other contexts. What makes an intelligent human reader grasp this meaning so unhesitatingly is, in addition to all the other features that have been discussed by MT workers — Dostert[4)], e.g., lists no less than seven of what he calls areas of meaning determination, none of which, however, takes care of our simple example —, this intelligent reader's knowledge that the relative sizes of pens-qua-writing-utensils, toy boxes and pens-qua-play-pens are such that when someone writes under ordinary circumstances and in something like the given context, "The box was in the pen" (and the occurrence of this sentence in the mentioned paragraph tends to increase the confidence of the reader that the circumstances are ordinary, though the whole paragraph could, of course, still have formed part of a larger fairy tale, or of some dream story, etc.), he almost certainly refers to a play-pen and most certainly not to a writing pen. This knowledge stands at the disposal of the average human reader beyond a certain age, and the writer takes this into account. This knowledge does not stand at the disposal of the electronic computer and none of the dictionaries or programs for the elimination of polysemy puts this knowledge at its disposal.

Whenever I offered this argument before one of my colleagues working on MT, their first reaction was: "But why not envisage a system which will put this knowledge at the disposal of the translation machine?" Understandable as this reaction is, it is *very* easy to show its utter futility. What such a suggestion amounts to, if taken seriously, is the requirement that a translation machine should not only be supplied with a dictionary but also with a universal encyclopedia. This is, however, surely utterly chimerical and hardly deserves any further discussion. Since, however, the idea of a machine with encyclopedic knowledge has popped up also on other occasions, probably also during the present conference, let me add a few words on this topic. The number of facts we human beings know is, in a certain very pregnant sense, infinite. Knowing for instance, that at a certain moment there are exactly eight chairs in a certain room, we also know that there are more than five chairs, less than 9, 10, 11, 12, and so on ad infinitum, chairs in that room. If you so wish, we know all these additional facts by inferences which we are able to perform, at least in this particular case, instantaneously, and it is clear that they are not, in any serious sense, stored in our memory. Though one could envisage that a machine would be capable of performing the same inferences, there exists so far no serious proposal for a scheme that would make a machine perform such inferences in the same or similar circumstances under which an intelligent human being would perform them. Though a lot of thought should surely be given to the problems which could only be touched here very little, it would very definitely mean putting the horse before

the cart if practical MT would have to wait for their solution. These problems are clearly many degrees of order more difficult than the problem of establishing practical machine aids to translation. I believe that it is of decisive importance to get a clear view of this whole issue and hope that my remarks will contribute to its clarification.

I have no idea how often sentences of the mentioned kind, whose ambiguity is resolvable only on the basis of extra-linguistic knowledge which cannot be presumed to be at the disposal of a computer, occur on the average in the various types of documents in whose translation one might be interested. I am quite ready to assume that they would occur rather infrequently in certain scientific texts. I am ready to admit that none might occur on a whole page or even in some whole article. But so long as they will occur sometimes, a translation outfit that will claim that its output is of a quality comparable to that of a qualified human translator will have to use a post-editor. As soon as this is granted, the greatest obstacle to practical MT has been overcome, and the way is free for an unprejudiced discussion of the best human use of the human partner in the translation outfit.

Having shown, I hope, that FAHQMT is out of the question for the foreseeable future because of the existence of a large number of sentences, the determination of whose meaning, unambiguous for a human reader, is beyond the reach of machines[5)], let me now discuss this issue of reduction of semantical ambiguity a little further. There exist in the main two methods of reducing semantical ambiguity. One is the use of idioglossaries, the other is the already mentioned method of utilizing the immediate linguistic environment of the word which is ambiguous in isolation. Though some doubts have been raised on occasion as to the validity of the first of these methods, I do not know of any serious attempt to put its validity to test. At this point I would only like to impress you with the vital necessity of performing this test before an MT method based upon its utilization is claimed to yield high-quality translations, even in collaboration with a post-editor. It is just the great effectiveness of the use of idioglossaries in general which is apt to yield disastrously wrong translations on occasion without giving the post-editor even a chance to correct these mistakes. It is just because a certain Russian word in a chemical paper will almost always have a certain specific English rendering that the danger is so great that in those exceptional cases where this word, for some reason or other, will have a different meaning, this exception will not be taken into account, yielding a meaningful but wrong translation.

In regard to the second method, the situation is even worse, and has lately become even more confused through the use of certain slogan terms like 'thesaurus' in this connection[6)]. It is doubtless true that consideration of the immediate linguistic neighborhood of a given ambiguous word is a very powerful method, but it is again necessary to realize its limitations. I am not talking any more about those limitations which I pointed out through the use of my sample sentence. I am now talking rather about the fact that many MT workers seem to underestimate the importance of those cases of reduction of polysemy which cannot be obtained by looking at the immediate neighborhood, and even more so about the fact that partial successes in this direction have led many people to underestimate the depth of the remaining gap. Let me state rather dogmatically that there exists at this moment no method of reducing the polysemy of the, say, twenty words of an average Russian sentence in a scientific article below a remainder of, I would estimate, at least five or six words with multiple English

renderings, which would not seriously endanger the quality of the machine output. It is looking at the quantities involved which creates a distorted picture with many people. Many tend to believe that by reducing the number of initially possible renderings of a twenty word Russian sentence from a million (which is the approximate number resulting from the assumption that each of the twenty Russian words has two renderings on the average) to some eighty (which would be the number of renderings on the assumption that sixteen words are uniquely rendered and four have three renderings apiece, forgetting now about all the other aspects such as change of word order, etc.) the main bulk of this kind of work has been achieved, the remainder requiring only some slight additional effort. We have before us another case of what, in a superficially different but intrinsically very similar situation, has been called the "80 per cent fallacy".[7)] The remaining 20 per cent will require not one quarter of the effort spent for the first 80 per cent, but many, many times this effort, with a few per cent remaining beyond the reach of every conceivable effort.

NOTES

1) See "Decision procedures for structure in natural languages", Logique et Analyse, N.S., 2:19-29 (1959), which is a revised version of a talk given before the Colloque International de Logique, Louvain, September, 1958; the talk given before the Second International Congress on Cybernetics, Namur, September 1958, on "Some linguistic obstacles to machine translation" will be published in the Proceedings of this Congress.

2) See, especially, Chomsky, N., Syntactical structures, 's-Gravenhage,. Mouton & Co., 1957.

3) This memorandum is reprinted as Chapter 1 of Locke, W.N. and Booth, A.D., eds., Machine translation of languages, New York, John Wiley & Sons, 1955. The quoted passage appears there on page 21. For Reifler's and Kaplan's studies, see ibid., p. 227.

4) See Dostert, L.E., "The Georgetown-I.B.M. Experiment" ibid., Chapter 8, especially pp. 129 ff.

5) I am afraid, therefore, that Weaver's hopes, reuttered in 1955 in his foreword to op.cit., that forthcoming research on logical syntax and semantics will make it possible for a computer to produce an output that would require no more than polishing up by a post-editor — see ibid., p. VII — will not materialize. I am singling out Weaver just because his misjudgment cannot be explained as being the result of vested interests.

6) Notice, e.g., that the very same — fictitious! — thesaurus approach that would correctly render 'pen' by 'plume' in the sentence 'The pen was in the inkstand' would incorrectly render 'pen' by 'plume' in the sentence 'The inkstand was in the pen'.

7) See Bull, W.E., Africa, Ch. and Teichroew, D., "Some problems of the 'word'", ibid., Chapter 5, p. 98.