Time flies like an arrow

A new enthusiasm for the potential of machine translation is currently rising out of the debris of the collapse of interest in the subject of a decade ago. This article—the first of two on the topic—describes the new analytical approaches that attempt to translate ambiguous phrases such as the title above

Dr Yorick Wilks
is reader in theoretical
linguistics, the
University of Essex

There is an ancient Chinese curse that dooms its recipients to live in an interesting age, and by those standards machine translation workers are at present having a bad time. Things are interesting at the moment because there are conflicting claims in the air about how to do machine translation, and whether it can be, or indeed has already been, done. Such a situation is unstable, and we may confidently expect some kind of outcome—always cheering for the empiricist—in the near future. What has happened is threefold. First, the early "brute force" methods of machine translation that were thought have become extinct more than 10 years ago in 1966, to have surfaced again, like some coelacanth from the deep. These systems are now being sold under such trade-names as LOGOS, XYZYX, SMART and SYSTRAN; and the last, and best known, is now undergoing extensive testing in Paris and Luxembourg.

Secondly, some large-scale, more elegant and theoretically based, machine translation projects that continued—usually in universities—after the 1966 collapse are now actually being used, though sometimes on a scale smaller an that originally envisaged. METEO, for example, in Montreal, which was to have translated official documents from English to French is now in use for the translation of the more limited world of TV weather reports, which have to be given in both languages all over Canada. Thirdly, workers on natural language in the field known as artificial intelligence (AI) have begun to make distinct claims about the need for their methods if there is ever to be general and high quality machine translation. Small pilot systems illustrating their claims have been pro-grammed, but their role in contemporary discussion is mainly that of a theoretical goad.

But these are not just three complementary approaches, for they seem to be making different claims, and, unless we take the easy way out and define some success level of machine translation appropriate to each of the enter-prises, it seems they cannot all be right, and that we may hope for some resolution before too long.

What we now have is four generations of research on machine translation: the original efforts of 1957-65, on which millions of dollars were spent, plus the three types of project now surviving, and indeed competing. The key to their relationship can be found in their different responses to a famous critique of machine translation by Yehoshua Bar-Hillel. He updated it at intervals, but it came down to one essential point: machine translation *is*
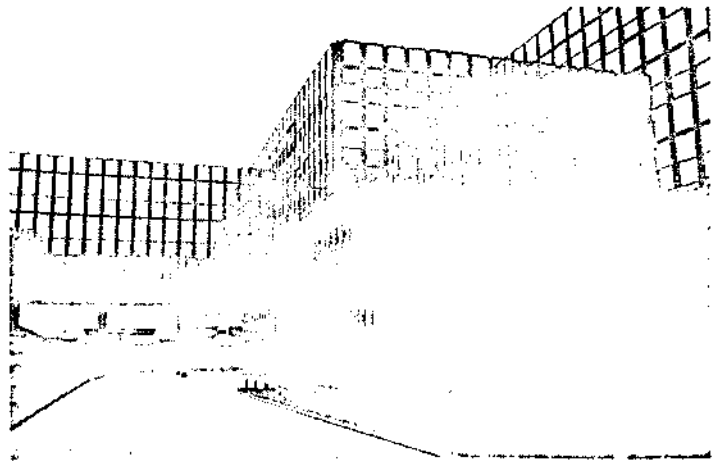
The Jean Monnet Centre, Luxembourg, where the SYSTRAN system is working

not only practically, but theoretically, impossible.

"Expert human translators use their background knowledge, mostly subconsciously, in order to resolve syntactical and semantical ambiguities which machines will either have to leave unresolved, or resolve by some 'mechanical' rule which will ever so often result in a wrong translation. The perhaps simplest illustration of a syntactical ambiguity which is unresolvable by a machine except by arbitrary or *ad hoc* rules is provided by a sentence, say '. . . slow neutrons and protons . . .' whereas, in general, though by no means always, the human reader will have no difficulty in resolving the ambiguity through utilisation of his background knowledge, no counterpart of which could possibly stand at the disposal of computers."

Problems of verbs

The immediate historical context of Bar-Hillel's argument was the performance of early syntax analysers which, according to legend, were capable of producing upwards of 10 grammatical parsings of sentences like "Time flies like an arrow". With respect to standard dictionary information, any of the first three words in the sentence could be taken as a possible verb. To see "time" as the verb, think of the sentence as a command with the accent on the, first word; to see "like" as the verb, think of the sentence as expressing the tastes of a certain kind of fly, and so on.

The standard reaction to such syntactic results was to argue that this only showed the need for linguistic semantics, so as to reduce the "readings" in such cases to the appropriate one. Bar-Hillel's response was to argue that it was not a matter of semantic additions at all, but of the, for him, unformalisable world of human knowledge.

The contrast can be seen by looking at our everyday understanding of so simple a sentence as "He paddled down the river in a canoe". The standard machine parser, working only with grammatical information, will not be able to decide whether the clause "in a canoe" attaches to "paddled" or "river". The first reading, the correct one of course, tells you *how* he went down the river. The second implies that he went down a river that happened to be inside a canoe—the same structure that *would* be appropriate for "He paddled down the river in an unexplored province of Brazil". The purely syntactic parser has no way of distinguishing these two possible "readings" of the sentence and, furthermore, there is a difference of opinion as to how the information that would resolve the problem should be described. Those who take a more "linguistic semantics" view would say that it is part of the meaning of "canoe" that those objects go in rivers and not *vice versa;* whereas those of an AI persuasion would be more likely to say that it is merely a fact about our world that canoes are in rivers. At bottom, there is probably no clear philosophical distinction between these views, but they do lead to different practical results when attempts are made to formalise and program such information.

It is interesting to notice that the reactions of Bar-Hillel and AI workers like Marvin Minsky were in part the same: Minsky argued that machine translation required the formalisation of human knowledge, programmed in a system that could be said to understand; or, as Bar-Hillel reviewed the situation in 1971: "It is now almost generally agreed upon that high-quality machine translation is possible only when the text to be translated has been understood, in an appropriate sense, by the translating mechanism."

What Minsky and Bar-Hillel disagreed about was what followed: Bar-Hillel thought that the absolute impossibility of high-quality machine translation had been demonstrated, whereas Minsky believed that the task had now been defined, and the job of AI was to get on with it.

The contrast is clear between the views of Bar-Hillel and Minsky on one hand, and the views of linguists on the other: Noam Chomsky's generative theories are also, in a clear sense, a reaction to the failure of the early machine translation work, in that they state the case, with great force, for a solid theory of the syntax of natural languages as a precondition for any advance with machines and language. Jerrold Katz and Jerry Fodor's semantics, joined to a Chomsky grammar, represent, as it were, the linguistic analogue to those in machine parsing who thought that purely semantic information would be enough to resolve the multiple analyses of the notorious "Time flies like an arrow". Later linguists broke from the Chomskian paradigm by arguing that Katz and Fodor's rigid exclusion of commonsense human knowledge from a linguistic system was inadequate, and that many forms of pragmatic knowledge would be required in a full linguistic system.

The attempt by AI research to respond to Bar-Hillel's challenge is of a different sort. It is an attempt not only to admit from the beginning the need for "knowledge structures" in an understanding system, but also to formulate theories and systems containing processes for the manipulation of that knowledge. "Processes" here is not to be taken to mean merely programming a computer to carry out a task, for many interesting AI systems have either not been programmed at all or made to do only partial demonstrations. The word "process" means that a theory of understanding should be stated in a symbol-processing manner, one in which most linguistic theories are not. This is a contentious position, because generative grammar has also been in some sense a description of a process since the earliest descriptions of transformational theory. The AI case is that it never quite comes up to scratch in processing terms. The nature of this dispute can be seen from such recent work as that of Joan Bresnan where an attempt is made to present transformational grammar at the highest level on an unfamiliar (to linguists) process-oriented manner.

The METEO system represents the survival of the linguistic tradition in machine translation work: with its claim that machine translation system based on a linguistic theory is sufficient, and that whatever knowledge is required for translation can be expressed within a grammar-based system containing a semantic component.

The contrast with the resurrected "brute force" methods should now be clearer. These approaches have just ignored the challenge of Bar-Hillel, as well as the earlier one from linguistics, for a theoretically motivated syntax and semantics. They have simply kept on going, and the chief assumption behind work like SYSTRAN is that the main fault of early machine translation was inadequate machines and software, not theory.

Striking demonstrations have been given of the SYSTRAN systems: at the University of Zurich on 12 June, 1975, before Swiss professors and military officers, the system successfully translated 30 000 words of unseen text from Russian to English on a machine of the university's, not SYSTRAN's, choosing. This is not conclusive perhaps, though it is far more solid than the public demonstrations that used to be given of first generation machine translation. Moreover, the details of the SYSTRAN system are understandably not available, since it is a commercial rather than a research enterprise. But there can be no doubt that it poses a considerable challenge to the linguists and the AI theorists both of whom claim, in their different ways, that some kind of higher level theory is essential for any reasonably high-quality machine translation.

Unlike Britain, where machine translation was simply written off as a research enterprise after the 1966 collapse, the US government funding agencies keep reviewing the possibilities of starting it up again, suitably armed with some new theory. So Winfred Lehmann and Rolf Stachowitz in 1971 wrote a report on whether or not generative lin-

guistics would justify a new start; and, more recently, Jim Mathias and Dave Hays have conducted a similar seminar, largely devoted to the question as to whether AI theories of language would produce the goods. But what is an AI theory of language, and how might it help machine translation?

AI has been concerned, for some 25 years now, with the problems of human intelligence seen from a particular point of view: what would it be like to program a computer to perform intelligent tasks that we do without even thinking about them; such as seeing and understanding what we see, understanding language, and inferring from what we understand? Some choose to investigate machine performance of tasks, like chess playing, that even humans find difficult, but the "unconscious tasks" remain the heart of AI.

As applied to the field of natural language understanding this has meant constructing elementary programs to carry out written commands, translate into another language, make inferences, answer questions, or simply carry on a dialogue—all of which are presented as written responses at a teletype or video screen.

Acts of translation

As can be seen, machine translation is by no means the typical AI language program, but no difference of principle arises between the different sorts of task, especially if we accept a slogan like George Steiner's that, in some sense, all acts of understanding are acts of translation.

What almost all AI language programs have in common —though they differ widely over other assumptions—is strong emphasis on the role of knowledge in understanding, and on the presentation of a theory as a possible process. In some programs—like a well known one constructed in 1971 by Terry Winograd—this last assumption leads to writing the syntactic analysis part of the program in a special "grammar programming language" PROGRAMMAR, rather than as the normal battery of grammar rules like S → NP + VP. This rule appears in all grammars and simply means that a noun phrase (NP) followed by a verb phrase (VP) is a well-formed sentence (S). In Winograd's system that grammar rule exists only as a tiny program in PROGRAMMAR.

Winograd's program accepted dialogue and commands about a miniature world consisting only of a few blocks and a box, which it could appear to move about on the video screen. He wanted to show the role of knowledge of this microworld of blocks as a tool for resolving syntactic ambiguities in input to the system. So, for example, when it saw the sentence "Put the pyramid on the block in the box", it would immediately resolve the surface syntactic ambiguity of the command: that is, does it refer to a particular pyramid (on a block) to be picked up, or to a particular place to put it (on the block in the box), according to whether there actually was a block under a pyramid, or already in the box, in the small blocks scene that it understood.

Winograd's program could be called pure AI, in that it was motivated by classic problems of AI: plans (how to pick up the blocks) and theorem proving (how to show which is under the pyramid at a given moment), rather than being motivated by problems left over from the 1966 failure of machine translation, such as word-sense ambiguity, and correctly referring pronouns in discourse. Another group of AI language programs, the work of Eugene Charniak, Roger Schank and myself was directed more at those questions: at meaning representation, and the use of inference rules, not about microworlds of blocks, but about the more general world in which we live.

Consider a simple sentence like "The soldiers fired at the women and I saw several fall", where we may be sure that any native speaker of English will understand that

sentence (out of any further context, which may change matters, so let us leave it to one side) in such a way that "several" refers to the women and not to the soldiers. That cannot be done on any simple semantic (or syntactic) grounds since both soldiers and women are capable of falling. Correct reference of the pronoun "several"—and this might be vital in translation into a language where "soldiers" and "women" had different genders, for example—must almost certainly be done using some general inference rules like "If animate things have an object projected at them, they may well drop downwards". If the reader finds that implausible, he should ask himself just how he refers the pronoun in that sentence.

The type of knowledge expressed in that rule is what one might call partial—it is an inference that is not always true. It is a kind of knowledge that has no place in the very limited Winograd blocks world, but is central to the understanding capacities of the Gene Charniak, Roger Schank and Wilks systems. The three systems differ strongly in other respects: for example, the Schank and Wilks systems emphasise knowledge that can be expressed in very general terms, like the inference rule above, and develop notations of semantic primitives (actions like CAUSE, and CHANGE; entitles like THING, and MAN for human being,) in order to express this. In Charniak's systems, on the other hand, the knowledge is more specific to certain topics.

Machine translation has traditionally been much preoccupied with the problem of finding the topic in a text: in the "Time flies like an arrow" example, we would have the correct reading if we could find out, from wider context, that the sentence is about time, and not about flies or liking. The semantic system of Charniak tried to detect topic by specific cues, while the Schank and Wilks systems did so by general rules ranging over semantic representations expressed in primitives. In the Winograd system, on the other hand, topic can never be a problem because it is always the blocks world!

There is no doubt that AI systems can be brought to bear upon the problems of machine translation: my system has actually translated English into French and resolved word-sense and pronoun ambiguities that could only be resolved with the aid of the sort of partial knowledge used in the soldiers and women example. There is enough capacity in such systems to express knowledge about protons and neutrons so as to have no difficulty with Bar-Hillel's phrase "slow neutrons and protons". If he were to protest that it was ad *hoc* for the system to code *only* one of those entities, say, as being potentially slow, then one could reply by asking how he could know that humans do not understand this with precisely such a coding of knowledge.

But much may depend on one's choice of examples: it is not clear that the difficulty has been eased by these AI systems for old favourites like Time Flying. The partial knowledge systems I described might well know that things that flew were normally birds or planes, rather than time, and so they would have no reason to pick out the correct reading on such grounds. Given that flies can indeed be timed, such systems might decide that the "imperative reading" was the one most suited to the general knowledge about the world with which they had been programmed. This is a melancholy conclusion, because it suggests that our competence with such examples can only be credited to an ability to read them off a list of prestored cliches, together with the interpretation "we feel as if time moves quickly". This would be a sad conclusion for all theoretically motivated work, and an awful fate for a long cherished example!

In a second article (to be published shortly) I shall describe a recent shift in the attention of AI work on language—an attempt to construct and program yet more complex knowledge structures—one that might help even with this type of difficult example for machine understanding of natural language. □