LINGUISTICS

Machine Translation

Scientists will settle for less than best as engineers grapple with language arts.

"Times flies like an arrow," the punched tape read as it was fed into one of the latest translating machines. A brief second later, out came the translation—a phrase in Russian meaning "Time flies like to eat arrows."

In the United States, some $20 million spent by the Government on research and hardware development, including the work of hundreds of linguists and engineers, went into the forerunners of that machine. In nearly 30 foreign countries, too, quality machine translation has been a goal almost feverishly aimed at for decade. The reason is plain: the number of published words—especially in the sciences—is already astronomical and increasing at a near-vertical curve.

Many American scientists recall grimly that a Russian technical journal had strongly hinted at the feat of Sputnik I several months before the launching: it was not discovered in this country until an American scientist who knew Russian came across the article.

In spite of the years of effort, imperfect translations are still the rule. In fact, it wasn't until late in the 1950's nearly 10 years after the first suggestion that machine translation might be possible, that there was a general realization that the obstacle to good machine translation was not simply one of engineering. After experimentation, scientists have turned to language study.

"The people who tried the early projects were looking for quick results," according to Eugene Pronko of the National Science Foundation. "And they got them: quick and dirty. History has shown that the barriers are linguistic. It's made us back to more fundamental investigation into language."

As a matter of fact, at first thought, it does seem like quick results should be easy to come by: just feed the machine a translation of each word and then set it to work. But the story turned out to be complicated.

The problem can be seen in the sample, "Time flies like an arrow." There are at least two big ambiguities in this sentence. One, the problem with "like," shows up in the translation cited, in which the machine took the word to mean "enjoy" rather than "as." The other lies with the separate meanings of each of the first two words: what if the machine had read it not in the usual way but as an order to "clock some winged insects?"

"After trying word-by-word process-ing, what linguists found out," says M.I.T.-trained Prof. Paul Postal of Queens College in New York, "is that perfect translation requires a perfect linguistic theory and all the other knowledge a human translator has. A language user is always subconsciously relying on his knowledge. That's why if you want a mathematics article translated from Russian you're best off with a mathematician who knows Russian."

As a result, says International Business Machines Corporation of its team of linguists at the Yorktown Heights (N.Y.) laboratories: "We are directing our efforts toward a deeper understanding of the structure of language."

At Bell Telephone Laboratories, Dr. Lee B. McMahon recently developed a new form of English called FASE (Fundamentally Analyzable Simplified English), which may provide for more accurate translations by machines. FASE reads just like English, but all syntactical problems are removed. That is, the arrangement of the sentence is made clear, so that the usual sequence of subject-verb-object is followed. "Time flies," for example, would always be read in the expected way, for "time" in FASE is always a noun at the beginning of a sentence. Were we to mean "clock some winged insects," we should write in FASE, "Determine the speed of flies," which a machine would understand, since "determine" can only be a verb in the new language.

As it stands now, although machines can translate as many as 20,000 words an hour, their input must be specially prepared (translated into the new FASE, for example) and the output must be "post-edited"—all by humans. De-bugging the machine's output, in particular, is painfully time-consuming and more straining for the human post-editor than is straight human translation.

Virtually everyone working on machine translations is agreed that translation of human quality and economy from a machine is a thing for the distant future, if it is possible at all. What they are looking for is an inexpensive process that can produce just-readable translations of scientific writings; for it is the scientist—whose language is relatively straightforward and less loaded with ambiguous terms than, say, the language of a poet—who is clamoring for translations. And on the way to that goal, we have learned how little we know about the language we learn and use so easily.