FORUM ON MACHINE TRANSLATION

Machine Translation will not Work

Martin Kay
Xerox Palo Alto Research Center
3333 Coyote Hill Road
Palo Alto, CA 94304

PANELIST STATEMENT

Large expenditures on fundamental scientific research are usually limited to the hard sciences. It is therefore entirely reasonable to suppose that, if large sums of money are spent on machine translations, it will be with the clear expectation that what is being purchased is principally development and engineering, and that the result will contribute substantially to the solution of some pressing problem.

Anyone who accepts large (or small) sums on this understanding is either technically naive or dangerously cynical. It may certainly be that

1. machine translation could provide a valuable framework for fundamental research;
2. texts in highly restricted subsets of natural language could be devised for particular puposes and texts in translated automatically;
3. computers have an important role to fill in making translations;
4. translations of extremely low quality may be acceptible on occasions.

However,

1. the fundamental research is so far from applicability,
2. the language subsets are so restricted,
3. the useful computer technologies are so different from machine translation,
4. the quality of the translations that can be produced of natural texts by automatic means is so low, and
5. the occasions on which those translations could be useful are so rare,

that the use of the term in these cases can only result in confusion if not deception.

A determined attempt was made to bring machine translation to the point of usability in the sixties. It has become fashionable to deride these as "first generation" systems and to refer to what is being done now as belonging to the second or third generation. It should surely be possible for those who think that the newer systems can succeed where the earlier ones failed, to point to problems that have been solved since the sixties that are so crucial as substantially to change our assessment of what can be achieved. We know a good deal more about programming techniques and have larger machines to work with; we have more elegant theories of syntax and what modern linguists are pleased to call semantics; and there has been some exploratory work on anaphora. But, we still have little idea how to translate into a closely related language like French or German, English sentences containing such words as "he", "she", "it", "not", "and", and "of". Furthermore, such work as has been done on these problems has been studiously ignored by all those currently involved in developing systems.

Unfortunately, the sums that are being spent on MT in Europe and Japan are large enough to make virtually inevitable the production of a second ALPAC report sometime in the next few years. This will inevitably have a devastating effect on the whole field of computational linguistics, everywhere in the world. The report will be the more devastating for the fact that much of the money has in fact been spent frivolously, and much of the work has been incompetent, even by today's limited standards.