

Foreword

Machine translation was a matter of serious speculation long before there were computers to apply to it; it was one of the first major problems to which digital computers were turned; and it has been a subject of lively, sometimes acrimonious, debate every since. Machine translation has claimed attention from some of the keenest minds in linguistics, philosophy, computer science, and mathematics. At the same time it has always attracted the lunatic fringe, and continues to do so today.

The fascination with machine translation doubtless has many sources. No one who reads the King James Bible, or FitzGerald's *Rubaiyat of Omar Khayyam*, or any of the dozen or so translations that are made every month of *Scientific American* or *Geo* magazine, or the facing pages of the magazines the Canadian airlines give their clients, can retain any doubt that translation is one of the highest accomplishments of human art. It is comparable in many ways to the creation of an original literary work. To capture it in a machine would therefore be to capture some essential part of the human spirit, thereby coming to understand its mysteries. But just because there is so much of the human spirit in translation, many reject out of hand any possibility that it could ever be done by a machine. There is nothing that a person could know, or feel, or dream, that could not be crucial for getting a good translation of some text or other. To be a translator, therefore, one cannot just have some parts of humanity; one must be a complete human being.

Many scientists believe that there is insight to be gained from studying machine translation even when it is applied to very pedestrian kinds of text with the expectation of producing results of a quite pedestrian kind. This is because translation is a task that obviously exercises every kind of linguistic ability, except those involved with the production and perception of speech, without requiring any other contact with the world. For example, a scientist interested in how people

collect words to make up sentences, and sentences to make up texts, might seek to test his theories in a computer program that produced sentences and texts. Such a scientist would face the problem of how to provide his program with something that it wanted to say. Programs do not generally have needs or desires that they might expect to fulfill by talking, or perceptions of the world that might strike them as worthy of comment. The investigator would therefore be faced with the logically prior problem of how to provide his machine with an urge to talk. One way to do this that avoids the problems of perception, and most of the problems of motivation, is to have the program simply say what it is told to say. This proposal clearly runs the risk of trivializing the problem. The solution is to provide the stimulus in a different language from the one in which the response must be given. Much of the motivation for studying machine translation has been very different, coming from the perceived need to produce more translations, to produce them faster, and to produce them at lower cost. More and more of the world's commerce is conducted across national, and therefore linguistic, boundaries so that letters, agreements, contracts, manuals, and so forth, must be produced in several languages. Modern technology develops and changes faster than ever before, and the accompanying texts must therefore be replaced correspondingly sooner. Furthermore, advanced technology, such as cars, computers, and household appliances, is no longer available only to people with advanced education, so that the texts that are needed to operate and maintain these products can no longer be produced only in English. In the countries of the European Communities, in Canada, in many developing countries, and doubtless soon in the Soviet Union, translations are required for legal or nationalistic reasons even when a single version might fill the primary need.

In recent years, this perceived need for more translations than human translators can produce has led to a great increase in activity in the field, especially in Europe and Japan. To be sure, the aims of this work do not encompass anything comparable to the King James Bible or the *Rubaiyat of Omar Khayyam*. They do not even include the *Scientific American* or even advertising copy. In the near term, practical machine translation is intended to be applied to two kinds of material. The first is material that covers such a narrow subject matter, and in such a routine fashion, as to require little on the part of the translator that could really count as understanding. The second is material that will be read by people seeking only a rough idea of what is being said, so that an extremely rough translation will be adequate. Maintenance manuals for machines belong in the first category and the second finds its customers in companies with major competitors abroad and in government agencies with unlisted addresses.

Only under very special circumstances are the readers of translated texts allowed to see what emerges from the translation machine and, if they did they would surely be surprised at how different it is from what they actually get. This is because it has almost invariably undergone a process called 'post-editing' by a human translator. This, the defenders of the machine are quick to point out, is no different from what happens when translations are made in the traditional manner; an initial translation is checked by a 'reviser', who often changes the first version quite extensively. But critics of machine translation, of whom there are many, are more inclined to credit the translation to the post-editor, and only some

occasional hints to the machine. Certainly, even when applied to texts on very limited subjects by authors with no literary pretensions, machine translation still generally produces results that would not be acceptable from a human translator under any circumstances. Just what would improve the quality of the result, and even whether any substantial improvement is possible, are open questions. Indeed, it is an open question whether the great investment that has been made in the enterprise since the first systems were put to use in the 1960s has resulted in any real improvement.

As an observer of machine translation, John Hutchins has, in several other publications, provided many readers with their first introduction to the subject. Harold Somers has stimulated the interest of generations of students in it at the University of Manchester Institute of Science and Technology which, largely as a result of his endeavours, has also become one of the world's foremost centres for research in machine translation. They clearly both have opinions on the open questions of machine translation, but they are not the subject of this book. The reader who seeks insight into the open questions, however, will do well to begin here. This is a technical book, in the sense of explaining the parts of machine translation systems, the principles behind the parts, and such relevant theory as is available from linguistics and computer science. It is a non-technical book in that it assumes no prior knowledge of these matters on the part of its readers. The first part of the book gives all the background necessary for the remarkably detailed and insightful descriptions of several representative systems that make up the second part. There must be few people, even among those who are professionally engaged in this work, who would not find information in this second part that was new and surprising to them.

Martin Kay

