Why take the machine's word for it?

by John Hutchins (University of East Anglia, Norwich)

The computer has been associated with natural language translation since it first appeared in the mid-1940s, but the early enthusiasm was followed by a long period of disillusion. In the last decade, however, there has been a revival with the appearance of many computer-based aids for translators and microcomputer translation systems. The question is no longer whether computers can translate or not, but how well they translate and how much help they offer to translators. The aim is no longer the emulation of the best human translators but the more realistic goal of making the translator's work more productive.

The early systems were designed for mainframe computers and for batch processing. The results were crude and required substantial revision or 'post-editing'. However, they were the basis of later more successful systems. The Pan-American Health Organization (Washington), for example, has developed in-house systems for translating between Spanish and English, with steadily improving quality since 1976.

However, the best known example of a mainframe translation system is undoubtedly Systran. Its first installation was in 1970 by the US Air Force for translating Russian scientific documents into English - and it is still in operation today. Its most substantial implementation has been at the Commission of the European Community, translating since 1976 an increasing proportion of the rapidly mounting volumes of technical, economic and administrative documents which need to be translated among the languages of the Community. Much of the output is revised before leaving the translation service, but some is only lightly edited (or even sent out untouched) if recipients want to know only the gists of documents and can put up with stylistic and grammatical errors. Other large and long-standing users of Systran (now owned by the Gachot company in France) include General Motors of Canada, the Xerox Corporation, NATO, Dornier, Aerospatiale and the Nuclear Research Centre in Karlsruhe. In addition, Gachot has recently made Systran available on the French Minitel network, with potentially far-reaching implications.

The use of Systran by Xerox merits particular notice. Here the output is not post-edited, but the quality is achieved by writing the technical manuals in a controlled English which minimises ambiguities of vocabulary and structure. The costs of pre-editing are outweighed by the quality of the simultaneous translations into five languages. The Smart Corporation (New York) has taken the same approach in a series of tailor-made mainframe systems for a number of large organizations: Citicorp, Chase, Ford, and largest of all, the Canadian Ministry of Employment and Immigration. A computer 'text editor' ensures clear writing, good grammar and style, and above all consistent use of terminology, before submitting text for translation (almost as a 'by-product') into French and other languages.

At the other end of the scale, powerful microcomputers have enabled the development of multilingual word-processing software and of relatively cheap packages for computer-aided translation intended for the individual translator and for translation agencies. These are the commercial by-products of many years of basic research which has produced effective (but not perfect) methods of syntactic analysis, procedures for dealing with ambiguities and facilities for bilingual dictionary compilation. Some limitations can be overcome by restrictions to specific subject fields. Many systems incorporate online interaction with human operators in order to solve

problems involving context or 'real world' knowledge, e.g. if the program cannot distinguish between light meaning 'not heavy' and light meaning 'not dark'. Inevitably, post-editing is still needed, unless recipients are happy with rough versions.

First in the field in the early 1980s were ALPS and Weidner, the latter becoming the market leader with its MicroCAT systems available for a wide range of languages of commercial interest, with the Japanese-English versions being particularly successful. Other microcomputer systems have also appeared in the last few years: the PC-Translator (Linguistic Products of Houston, Texas), the 'translating word processor' from Globalink (Fairfax, Virginia), and the Tovna MTS (Tel Aviv and Washington). The main selling points for these systems are their cheapness and their portability (nearly all run on IBM PCs and compatibles); but customers should not expect good quality output, and should be prepared to devote a great deal of time developing and refining dictionaries for their specific needs. More sophisticated, and dearer, systems are available for more powerful computers giving generally better results. The Logos Corporation has marketed German-English and English-German systems - installed at a number of German companies with large translation requirements, e.g. the computer company Nixdorf - and, more recently, the METAL German-English system has appeared, developed by Siemens and the University of Texas. It is probably the most sophisticated computer translation system on the market, with good quality raw output and excellent facilities for dictionary creation and post-editing. Further METAL systems are expected for English to Spanish and Dutch to French.

However, the most vigorous development of computer-aided translation has occurred in Japan. There are now a large number of microcomputer interactive systems on the market, nearly all for English-Japanese or Japanese-English, stimulated by the huge demand for translations of English scientific information and of marketing and technical documentation into English. Many are linguistically limited and rely on considerable human assistance. Examples are systems from Fujitsu (ATLAS/I), Oki (PENSEE), Mitsubishi (MELTRAN), Sharp (DUET), Sanyo (SWP-7800). As in Europe there are also more sophisticated systems capable of producing translations of rather better quality: e.g. from Hitachi (HICATS), Fujitsu (ATLAS/II Japanese-English), NEC (PIVOT) and Toshiba (AS-TRANSAC). Japanese texts present particular problems: three scripts, no capitals and no spaces between words. Japanese-English systems are intended for Japanese users who can do the essential pre-editing; users with no knowledge of Japanese will therefore get poor results.

While new systems are regularly appearing (e.g. TRANSTAR for English-Chinese translation), research continues worldwide. Much of this is based at universities, e.g. Grenoble, Saarbrücken, Kyoto, Carnegie-Mellon (Pittsburgh) and UMIST, but advanced research is now undertaken by many commercial companies. Most of the Japanese computer manufacturers are involved; elsewhere, for example, IBM in both Europe and US, Philips in the Netherlands, and a Utrecht software company are supporting basic research and developing prototype systems. Governments support research in many countries and there has been considerable international cooperation, e.g. the well known Eurotra multilingual research project for the European Community and the multinational project involving Japan, China and Southeast Asian countries. Research on computer translation of natural language is necessarily long-term and often speculative; it includes explorations of new linguistic theories, of artificial intelligence and knowledge based systems (in the view of many offering the best prospects for overall quality improvement), of statistical and pattern matching approaches, of computer advances in connectionism and parallel processing, and also speech processing and computer-based telephone translation (e.g. by British Telecom and by the Japanese ATR group.)

Despite the growing demand for translations which the profession cannot satisfy, there has been a relatively slow acceptance of computer-aided translation - with concomitant financial problems for some suppliers (e.g. Weidner). The reason is partly the low quality of the unedited output of cheaper microcomputer systems and the still doubtful cost-effectiveness of some more expensive systems, and partly the understandable caution of translators. Although higher quality will certainly result from current and future research it is nevertheless unlikely that computers will ever perform as well as the best human translators. Both in the short and the long term, translators need not fear for their jobs. Computers are aids to translation and not replacements of translators.