4.   MT UK: Britain and machine translation

Veronica Lawson

Associate Editor
COMPUTERS AND TRANSLATION

Manufacturers of computers and business equipment now rank high among the world's publishing houses, with IBM said to be the biggest publisher anywhere. The reason is the massive documentation needed for complex modern equipment - often 5000 pages per product per language. Exporting tends to mean an immense translating load, to be completed very fast. For speed and volume, industry is turning increasingly to computer-aided, or machine, translation.

Machine translation (MT) has come far since a notorious report in 1966 ended US government backing for research in the field. What with international marketing, increased political co-operation and the huge growth in scientific and technical information since the 1950s, there has been a "translation explosion" which shows no sign of abating.

So MT continued to develop, and it has long been in practical use in North America, Europe, the Soviet Union and the Far East. Research, too, is flourishing. Japan sees MT as an important part of its research into 5th generation computers. European research includes the ambitious Eurotra project, on which universities in all EEC countries are working with the European Commission.

Britain is a good example of the change in attitudes. It has a number of MT users (among them Xerox, one of the world's most successful), and its researchers include two teams working on Eurotra. The UK has produced many key MT publications and events, including the annual *"Translating and the Computer"* conferences. Three of the four commercial MT systems marketed in Europe - Systran, Weidner and ALPS, all from small US companies - are well known in Britain. (The fourth, Logos, still concentrates on Germany.)

Systran, developed by Dr. Peter Toma at the end of the 1960s, is probably the most generally successful MT system. Here is a thumbnail sketch based on information from Ian Pigott, Systran project leader at the European Commission. Numerous language pairs exist, but each Systran is a "one-to-one" system, translating from one source language to one target language. It has two linguistic components, one dynamic (the translation program proper) and one largely static (the big, powerful dictionaries). The two components interact throughout the translation process, which falls broadly into three phases: analysis of the source language, transfer between the two languages, and synthesis of the target language.

Analysis is by far the hardest of these, for it must establish the grammatical function of every word and its relationship with other words. First the words are looked up in the dictionaries to see whether they belong to a set phrase (like *"a matter of course"*), and to find what part or parts of speech they may be. Scores of special homograph routines then work on the numerous words which can be more than one part of speech, like *"can"* (noun, verb, adjective). After this, five syntactic passes explore the sentence, breaking it up into clauses, to which grammatical functions can then be attached. This operation reveals the structure of the sentence, and the various words and phrases can now be labelled to show which part of the sentence they are. Analysis (or parsing) is now complete, and the program goes to the transfer dictionaries to find the corresponding words in the target language. Contextual dictionaries are used to remove ambiguities. Finally, the target words chosen are fed to the synthesis routines, which add appropriate endings and re-arrange the words into the right order for the target language.

Xerox translate 4 to 8 million words a year by Systran - about 80% of all their service documentation, and possibly as much as 25% of the world's current MT production. The documentation - maintenance and diagnostic procedures, technical data, training material - is written in a restricted English known as **Multinational Customised English**. At first, in 1978, this was very limiting, with only 3000 words and a number of writing rules, but now the dictionary is far larger and constraints are few. Xerox's Systran then translates into German, French, Spanish, Italian and Portuguese, producing output which can be post-edited in a fifth of the time it takes to make a manual translation.

Many of Xerox's most successful products are designed in Britain, at Rank Xerox Ltd. in Welwyn Garden City, the centre of technical competence for European-designed hardware. The documentation for these products - some 40% of Xerox documentation - is prepared in Welwyn. Writers work at wordprocessors connected to a minicomputer, producing text which is transmitted on-line to New York for translation by Systran on a mainframe. The raw output comes back to Welwyn for post-editing, still in electronic form. The post-editors are usually Xerox service personnel brought over temporarily from the target market.

Xerox launch new products multi-nationally, i.e. as near simultaneously as possible. The last technical change may be made only three months before, and yet within this brief "window" 2000 to 5000 pages of documentation must be translated into all the necessary languages. Worse still, major design changes at the last minute are becoming commoner as more machines are controlled by microprocessors. Replacing one of these may change many of the machine's characteristics overnight, including diagnostics and the message sets which flash up on displays. Whole sections of a manual must then be amended and re-translated.

The translation workload fluctuates too violently for an in-house language service to cope, and outside agencies would not have the technical expertise. If Xerox companies are to launch products on time, according to Mike Scott, business and systems manager at Welwyn, only MT will supply consistent translations at an economic price

Companies outside Xerox may be able to use their MT service in the future. Welwyn already offers a competitive printing service to the public, and recently won a much-coveted contract to print patent applications for the European Patent Office in Munich.

Systran translations are also becoming available through service bureaux set up by the European Commission, e.g. at the Kernforschungszentrum Karlsruhe. The UK Atomic Energy Authority is co-operating with KfK on Systran dictionary updates, with a view to transmitting machine-readable text for translation in due course.

The Commission bought Systran in 1976 and have improved it extensively. However, the EC's seven official languages make up 42 language pairs, and this will rise to 72 when Spain and Portugal join. Systran covers only a few of these, and does not reflect all recent developments in machine translation, linguistics, computers and artificial intelligence. The Commission therefore decided to initiate an advanced machine translation project, **Eurotra**.

Whereas Systran was conceived as a bilingual system, Eurotra is to be multilingual, translating between all EC languages. It is to be portable from one make of computer to another, highly modular, and extensible, so that new languages, subjects, even research can be incorporated. Analysis, transfer and synthesis will be kept separate, sharing only their interfaces, and the modules for the different languages will be created by independent groups in the various states, to encourage research and accommodate different approaches. The total R & D grant is 27 million European Currency Units (about $22 million).

Researchers from UMIST (University off Manchester Institute of Science & Technology) and the University of Essex assist with central planning in Luxembourg as well as working on the English language modules. As yet Eurotra has existed chiefly as a set of increasingly detailed technical specifications, but prototype software is now running in UMIST, and the computer industry is tendering for the environment software this autumn. A pilot system should be translating in late 1986, using a dictionary of 2500 entries in each language. By mid 1988 a pre-industrial prototype, ten times as big, should show whether Eurotra can be expanded for use in the Commission and elsewhere.

One of the companies expecting to tender for the Eurotra software is **International Computers Ltd.**, which has had an interest in MT for some years. ICL translates its manuals with the Weidner system, which it has adapted to its own networked microcomputers (the DRS or Distributed Resource System). ICL's Graham Hook says that they plan to market this version of Weidner in due course.

ICL are also involved in a joint MT venture with UMIST and Sheffield University, funded by the British government. The aim is to translate technical manuals (initially computer manuals) from English into Japanese, and a prototype is being demonstrated this autumn.

In 1985 **Weidner** founded a new company, Weidner Translation (Europe) Ltd., near Southampton. Preferring to say computer-aided translation (CAT) rather than MT, they supply both MacroCAT for minicomputers and MicroCAT for personal computers. MicroCAT runs on the IBM PC/XT or AT, and WTE also have a version for ITT's Xtra XP micro, which is claimed to be fully IBM-compatible, and faster and cheaper than either. The XP (for "extra power") is available through WTE. ITT themselves are Britain's oldest Weidner customer, having used the program since 1982 both for their own work and in bureau service.

Weidner, like Systran, is a batch system, particularly suitable for large volumes of fairly predictable material like manuals. For shorter, less standard texts both ICL and Xerox are considering ALPS, a less ambitious system which relies on human assistance during the translation process. ALPS also offers the simpler options of multilingual wordprocessing and selective or automatic dictionary lookup.

This autumn, for the first time, commercial MT is being used to teach British students. **Bradford University** has bought Weidner for their translation students, and **Coventry** language students will be learning from ALPS. Students, it seems, now expect machine translation in their courses, regarding it as an inevitable part of their working life.

8