

# Machine translation for Arabic

by Diana Darke

Successes have recently been scored by several companies offering Machine Translation (MT) software for European languages: users are increasing and some systems are running on micros. On the Arabic front, success to date has been more elusive, but not for lack of interest.

The formidable task of developing software for Arabic MT is such that only two organisations have so far tackled it seriously: the Weidner Corporation of the United States and the Systran Institut in Stuttgart. Both have concentrated on English to Arabic, as 90% of the demand is for that direction, and whilst Weidner have had discussions with a Tunisian company over Arabic to English, no firm development project is yet under way. Both systems need medium-sized minis: Weidner's runs on DEC's PDP-11/44 with 1 megabyte of memory, and the final cost of the system including the hardware can range from £100,000 to £250,000; Systran's English to Arabic system 803-A runs on an IBM 370/360 (or compatible hardware) with 600 Kbytes of memory.

The problems of writing MT software for any language are immense, but with Arabic these problems are on a truly daunting scale.

All the same problems as exist in Arabising computers are encountered again, but on top of these, the machine must actually "understand" the structure of the language. Arabic grammar is held to be among the most difficult in the world, and the sheer number and complexity of the rules makes programming nightmarish. The process is compounded by Arabic's root system on which all derivative meanings are based (as explained in my article on the Arabic language in the October issue of *Language Monthly*), and by the fact that only

consonants are written, leaving the reader to supply the vowels for himself from his knowledge of the grammar and the context. This involves a considerable process of deduction, as the meanings of words with identical spelling will change depending on the vowelising. Hence what in Arabic is written "MQDM" can mean giver/submitter, prefaced, forearm, Lieutenant Colonel, or the front part of a ship or aircraft, depending on the vowelising the reader supplies. For the computer to deduce these meanings from context is exceptionally difficult. On top of all this, there are ten possible derived forms of the verbs and the many patterns required just to form the plural of nouns.

To cope with these problems, Weidner has had to program each verb in the system with over 150 different basic patterns, each having 106 different inflectional possibilities. Nouns and adjectives are inflected on more than 250 basic patterns, each having 15 further possible inflections. When the translator wants to enter a new word in the system dictionary, Weidner have, thank God, designed the software so that he only needs to supply two basic forms for nouns and adjectives, and four basic forms for verbs. Once entered, the software carries out an analysis of the required forms and automatically assigns the appropriate inflection rule number.

Weidner are as yet the only company to have a commercially available English to Arabic system. Systran

have devoted considerable efforts to their system, but it is still not fully operational. The American company ALPS offers five language combinations from English, but only offers Arabic in WP and dictionary looking-up facilities. Logos are also reported to be working on a system, but are still some way from a working product.

With the problems and limitations of Arabic MT, it is not surprising that success in the sale of systems to date has been limited. Weidner have in fact recently moved back from direct marketing of the system to some extent, by selling the rights of its English to Arabic system to a firm in Riyadh called TAARIB, with whom they have a joint venture arrangement. Weidner claim that their system is being used by translation service bureaux in Saudi Arabia and in the United States. Two further installations were made in Oman and in Qatar, both in government departments. Both were sold via a company called Omnitrans, a Californian company, with whom it seems Weidner no longer has relations. As a result, these systems are not being supported with periodic updates, and the system in the Omani Palace Office is simply being used as an expensive plant stand.

According to its users, it was capable of translating sentences of the complexity of "the cat sat on the mat", but anything with sub-clauses came out so garbled that it took longer to post-edit than it would have done to write it out manually in the first place. The raw MT it produced was often so radically different from the original that if the post-editor was not on the alert, the sense of the final translation could be

the exact opposite of what was intended. Omnitrans were also said to be using the Weidner system to translate the Encyclopaedia Britannica into Arabic, using an Optical Character Recognition system to feed in the original.

Overselling the capability of systems seems to be a common failing of MT companies. Clients are told the systems will double or even quadruple the average translator's output, give consistent terminology and provide camera-ready copy. The Omani Palace Office was told their system would be 80% efficient, but this is not what they have found.

Remarkably, despite the shortcomings and the costs, interest in the possibility of Arabic MT remains high, especially in Saudi Arabia. Like the computer, it is regarded as a "technology pipeline". The theory is that if technical documents and new research papers can be speedily translated into Arabic, technology can be transferred without the need to import expensive experts and advisers. Accuracy in such a field would, however, be essential, and clearly it is here that Arabic MT still has a long way to go.