

Computer-aided translation on personal micro

The MicroCAT, a Weidner computer-aided translation system on IBM Personal Computer, was launched on June 12 in London at the IBM Computer User Show by the UK distributors, The Software Connection, of Fareham in Hampshire.

The launch marks a major step forward in making available to practising translators computer-aided translation technology, which up to now has not found wide application in Europe outside the European Communities and a tiny handful of large companies.

major theme

The Weidner Communications Corporation, of Northbrook, Illinois, United States, has for some years been in the forefront of marketing a CAT (computer-aided translation) system. This system has now been adapted for an enhanced IBM PC XT Personal Computer, and an agreement made with the Software Connection for its distribution. The Software Connection is a member of The Professional Connection group of companies, which produce, distribute and sell computer, word processing and microcomputer hardware and software, not only in the UK but also in European countries with the exception of France.

The special Weidner software offered is currently in the language directions English to French, English to Spanish, English to German, English to Portuguese, French to English, and Spanish to English.

Each language direction has a core dictionary of around 10,000 words and idioms. In addition there is a dictionary update facility which allows users to add or delete words or phrases at will, so making it possible to create word banks or bilingual glossaries specific to an industry, a company, or the needs of an individual translator. This offers considerable simplification in the translation of technical documents, scientific reports and parts lists, and allows terminological consistency.

Raw translation can be generated at rates of 1,600 words an hour upwards, and the translator's task is to concentrate on refining and editing the text which appears on the screen.

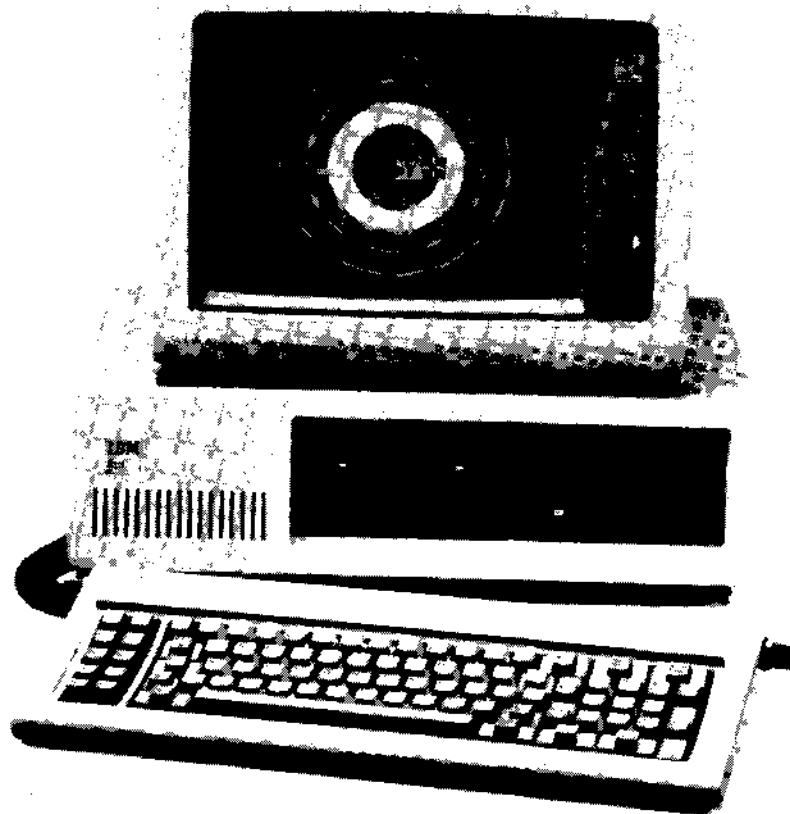
One suggestion made is that to derive maximum benefit from the time-saving, texts are organised in the afternoon for deferred translation and are ready for editing when the translator arrives for work the following day.

Editing as a post-translation exercise is done using a split screen facility, which enables a translator to

examine the source text and the raw translation in parallel.

sentences

The system analyses and translates by sentence, not by word. Verbs, idioms and phrases of up to nine words and a maximum of six alternative meanings can be dealt with, as



can homographs, such as the verb "conduct" and the noun "conduct". It also performs a syntactical analysis and re-orders the words to suit the target language.

The dictionary, claimed to be unlimited in size, is at the heart of the system.

Source documents can be entered into the MicroCAT by keyboarding, from a word processor, or by optical scanner. Translated documents can be produced as a camera-ready print-out or on tape.

It is claimed that translators trained on the system can edit between 600 and 1,000 words an hour, depending on the completeness of the dictionary, the style of writing, and their own proficiency. A maintenance agreement can be taken out as part of which updated versions of the software package are supplied as they appear.

options

No pre-editing is required, so the source material can be fed directly into the system. The system is menu driven, with access to a list of all the standard programs and options.

Among the options are Vocabulary Search, an important feature of MicroCAT. As the quality of raw translations is totally dependent on what the computer "knows", the vocabulary search option gets the computer to process the source text for words it does not recognise. It then provides a list of all such words for entry into the system's dictionary, if desired. The list can be sorted alphabetically, by frequency or occurrence, or by context.

Another option is Dictionary Update, which allows the translator to customise the system to his own unique vocabulary requirements.

four stages

The way the computer accomplishes raw translation can be broken down into four stages.

Pre-analysis is the process of breaking the source text into the smallest units which are meaningful to the computer. The first step is accomplished by the pre-processing routine which isolates sentences, inserts formatting markers, and

identifies typesetting commands. The dictionary module then morphologically analyses source words to remove inflections and locates the definition of the stem forms in the system dictionary. Inflection information is preserved for later use in the translation process.

analysis

The second stage is analysis. Here the computer again examines the small units it has identified and labelled, and begins to reconstruct them at the next highest level of meaning — the phrase. This is known as parsing. The compaction routine isolates word groups that act as a single unit, such as verb strings and idiomatic phrases. These word groups are then compacted into a single unit for the computer to deal with, and are further flagged with syntactic and semantic information for later analysis. Homographs are examined within their context to determine which meaning should be selected. Finally, the phrase structure analysis routine performs a left to right "parsing" of the source text, identifying progressively higher levels of phrase structure. Phrase and clause structure information generated by this analysis is preserved for subsequent reorganisation into the structure of the target language.

transfer

Transfer is the third phase, in which the structure of the source language is transformed to target language structure. The insertion routine inserts articles, prepositions etc., into the text as required by the target language. It also deletes redundant words from the source language. Words and phrases marked for re-ordering are moved to the appropriate positions in the target language sentence and the text is reformatted accordingly. The expansion routine transforms the phrases that were previously "compacted" into their corresponding target language structure.

synthesis

Synthesis is the final phase. Here the translated text receives its final

touches. Inflection of verbs, adjectives, determiners, nouns etc. is performed to provide "agreement" between sentence elements. Then the cosmetics routine makes the miscellaneous orthographic and other necessary low-level adjustments that are required in the target language output.

There is also a post-processor routine, which puts the machine translation into the format of the original text, restores typesetting commands, and performs special handling of untranslated words or phrases. The machine translation is then stored in the typed file awaiting instructions from the human translator.



John Newton

Mr John Newton, formerly with I.G.G. Electronics and Linguaphone, an Associate of the Institute of Linguists, has been appointed marketing manager of The Software Connection with special responsibility for microCAT.

What will it cost? our readers will be asking. Speaking by telephone to Language Monthly from the exhibition stand, John Newton pointed out that the pricing was a very complex business, depending on what you wanted. But the machinery, the IBM PC XT enhanced micro-computer CPU, the 10-megabyte hard disk version with 640 K RAM, with mono screen and keyboard, at present list prices, would cost you £5,244. The software would start at something like £8,000 for the first language direction, with prices reducing for each additional language direction acquired.