BSO launch their DLT prototype

MT system which resolves ambiguities by asking the author

The first working prototype of DLT, one of the more original machine translation (MT) systems now under development, was unveiled and demonstrated at an international press conference in Utrecht (Netherlands) on December 1.

DLT (Distributed Language Translation) is being developed by the Dutch software company BSO (*Bureau voor Systeemontwikkeling*), and work began on it in 1982. The December 1 presentation took place at BSO's Utrecht headquarters.

Toon Witkam

The DLT prototype demonstrated at the press conference is still only a laboratory version equipped with an initial vocabulary of some 2,000 English words, but the key ingredients of the future production model are there. Sentences typed in (from a technical handbook) in English are first converted to the intermediate version in Esperanto.

DLT is designed to translate informational, technical and business texts from and into various languages, over computer networks. A first commercial version of the DLT system is expected to be on the market by 1993, the press conference was told. This first version will translate from English to French; other languages will be added in the course of the 1990s.

The DLT approach is based on a network concept, with each terminal translating into or from a chosen language, but with the common links being provided by an "interlingua" (intermediate language). This interlingua remains unseen and unnoticed by the system's users, but a compact and standardised version of the text is translated into it and distributed to the receiver stations. Esperanto has been chosen as the basis for this interlingua.

If there is more than one possible translation, due to amiguity of relationship between words, or polysemy (more than one possible meaning) of a term, the more plausible interpretations are then separated from the unlikely ones by reference to a "knowledge bank" composed of some 80,000 "semantic atoms". In order to achieve this BSO's software makes use of techniques derived from research in the field of Artificial Intelligence (AI). After the interpretation selected as most likely has been confirmed by the operator, the intermediate product is translated into French.

Using AI techniques on present day hardware makes the process of translation rather slow. BSO expects this problem to disappear when the next generation of hardware, the so-called parallel computer, becomes available in the early 1990s. At the present stage it is more important for the developers to be able to study the "anatomy" of the system in detail.

Eckart Wintzen

The translation process can be followed on a bank of five graphic work stations like a slow-motion action replay, with the intermediate stages displayed in the form of tree-like sentence structures (see illustrations on next page).

Although the DLT project has reached the halfway mark in its allotted period of research and development, a great deal of work remains to be done to build a full-sized system. This includes increasing the size and scope of the dictionaries and knowledge bank. It emerged at the press conference that one third, or some six million Dutch guilders (equal to £1,800,000 or $1 million), of the funds available for the period 1982 to 1991 has now been spent, which means that the project is not only on schedule but has kept strictly to budget. Of the total cost, estimated at 18 million guilders, 50% is being contributed by the Netherlands Ministry of Economic Affairs from funds available for research on new technologies, while the remaining half is provided by BSO itself.

The presentation was opened by BSO managing director Eckart Wintzen, who said that DLT should not be regarded as a "translation machine" as such, but as an advanced software component. "We are not", he explained, "building a tool for translators, we are building a tool to communicate".

The system, he added, was not yet finished, but BSO had now reached the stage where they had shown it

could work. Although the prototype was operating with a small number of words and a limited grammar, it did show that it understands the syntax of the language.

John Hutchins, author of the standard work on machine translation*, gave the audience a general historical survey of MT, and illustrated the problems which automatic translation had to face. He drew a distinction between direct translation systems, where one language is rendered directly into another, and indirect systems, such as DLT and the Eurotra system being developed by the Commission of the European Communities, which uses an interlingua.

He also described the increase in Japanese MT activity in recent years. "In the very near future", he warned, "the Japanese are going to threaten to dominate the market".

The thinking behind DLT was explained by Toon Witkam, initiator and leader of the project.

Journalists watching one of the demonstrations

John Hutchins

It was generally agreed, he told the audience, that FAHQT (fully automatic high quality translation) would not be possible in the next 50 years. So a compromise was necessary. Some systems had gone for "quick and dirty" output, requiring a lot of cleaning-up, but BSO had decided that this was not the path they wished to follow. Other possibilities were to provide for pre-editing, post-editing or interactive editing.

It had also become clear that for economic reasons there had to be a

facility for adding other languages, and it was now generally agreed that an interlingua system was the best architecture to use.

DLT, he confirmed, was not intended to be a tool for translators, but for people who know their own language, and wanted to generate a foreign language version of their text. It called for some help from the user, who would often be the author.

Esperanto had been chosen for the interlingua because it was a "fully fledged" language with both grammar and dictionaries. Asked whether BSO had considered using a number-based interlingua, he said it would be unreasonable to ask programmers to work with numbers, which conveyed nothing to them and which would require constant reference. In conversation later Mr Witkam said that early in the project (cf report in *Language Monthly*, April 1985) they had anticipated having to modify the Esperanto base, but in fact few modifications had proved necessary.

Translation, he told the conference, required a combination of language knowledge on the one hand, and "knowledge of the world" on the other. In DLT the language part was performed by the syntactic method known as metataxis. The knowledge of the world, or context interpretation, was more of a problem — "This is what we have spent most of the six million on!", he added. However the work done would make it possible later to connect in many other languages.

Mr Witkam explained — and this was subsequently illustrated in the demonstrations — how the operator was asked in computer interactive dialogue to resolve ambiguities, by making "semantic choices". By this means the success of the system rested on "two pillars — artificial intelligence, and human intelligence".

However, as information was gained there would be a gradual growth in the number of choices where the system could rely on AI. The graph of BSO's projection of this trend is shown on the front cover of this issue of *Language Monthly*.

This approach, said Mr Witkam, was "the safe route to AI". The machine would learn from experience over many years.

Looking to the future, the next stage was to produce the second prototype, to increase in the scale of the dictionaries, which might involve partnerships, and to move from a laboratory project to a more industrial project, from research to development. Field tests were scheduled for 1991 on.

"Machine translation provides a very exciting race. We have acquired a good strategic position in this race. We will continue with this project, and we are open for international cooperation, Mr Witkam concluded.