EEC FINANCE PROJECT

Can Esperanto help Machine Translation?

The first results are now out of a pair of EEC-funded studies designed to find out whether Esperanto could be used as a pivot language for machine translation (MT).

The idea ultimately is to machine translate first from the source language into Esperanto, which, being synthetic, is simpler and supposed to be less ambiguous than a natural language, and then from this intermediate language or pivot into the target.

The European Communities had given 100 000 ECU to a Dutch and a Danish firm: the Dutch software house BSO (Buro voor Systeemontwikkeling, of Utrecht) for a linguistic feasibility study, and Christian Rovsing International A/S for a dataprocessing and market study. It had been suggested that this approach should be tried because the 1966 ALPAC report had found that fully automatic high-quality machine translation was impossible, so that a more modest goal was more realistic; no-one, it was suggested, had yet been successful in machine translation (a matter of definition, perhaps). Academics were felt to need an understandable pivot if they were to follow the machine translation process, and so it was sensible to use an international language like Esperanto with a simple structure. Lastly, "distributed language translation" - by microcomputers networked together by means of telecommunications or even videotext systems was both necessary and, it was believed, novel.

The BSO report is in fact entitled "Distributed Language Translation - a multilingual facility for videotext information networks". The 200-page report, a model of report writing by A.P.M. (Toon) Witkam, is in English with summaries in French, German, Dutch and Danish. It is "aimed at MT specialists, computer-minded translators and computational linguists, at the heads of translation departments in industry and government" and is also "intended to find its way to the desks of planners and decision-makers in the publishing world and the computer and consumer-electronics industries.

Most of the Report concerns Esperanto, its fitness for the purpose and how it needs to be improved. It acknowledges immediately that the critical issue was a linguistic one (the unambiguity of Esperanto), which is why three-quarters of the report is of a linguistic nature. Other chapters deal with background, a general view of the proposed system in dataprocessing terms, hardware and software implementation, and a technical schedule for the next development phase, assuming that further financial support is forthcoming.

Esperanto, it turns out, does have disadvantages, as professional linguists (even outside the MT field) predicted. In particular, it suffers from a general lack of technical vocabulary. What technical terms there are have not been standardised: different nationals invent different Esperanto terms. The report suggests that technical Esperanto should be standardised before work on the MT project begins, but it is not clear how this might be done.

Another problem is that the use of Esperanto as an intermediate language has been found on occasion to increase ambiguity. The translation from the source language into Esperanto creates ambiguities particularly in the case of noun groups, for example because although a source language may indicate the relationships between nouns clearly (by means of inflections or basic syntax), Esperanto does not. The report therefore suggests the addition of diacritical information - grammatical markers. It is notable that BSO themselves have not processed any text. However, they may well have received some of the results obtained in the Esperanto version of the Saarbrücken University's MT system SUSY by Dieter Mass, who, with other Eurotra researchers and three people from the Esperanto Academy, is thanked in the preface.

As so often, the examples used in the report are not taken from real life. Sentences such as "the witch made the prince a frog" and "they don't know how good meat tastes" may well illustrate points, but the level of language (register) is not one normally found in commercial translation. Not many of us can live by translating fairy tales - or even graffiti, as in "'My mother made me a homosexual.' - 'If I buy her the wool will she make me one too?'"

Both reports suggest as novelties the use of automatic editing programs, although such systems have been used both inside and outside MT for some time.

The Rovsing report contains much information on the use of microcomputers and networks. Its market study is much less helpful, suggesting as possible markets for MT not only videotext, but tourist guides, political and diplomatic texts for international organisations, film subtitles and dubbing, novels, plays and journalism, among others. It does however consider human translation as an alternative. The linguistic examples given by BSO have little or nothing to do with these markets.

It seems surprisingly , given the importance ascribed to videotext as a potential market for MT, that neither of the studies has looked properly at videotext or similar condensed information. In fact videotext was machine-translated in about 1980 to our knowledge - unsuccessfully because of its syntactic peculiarities and its great reliance on layout to convey meaning.

As the first concrete work on distributed language translation it is proposed to machine-translate from Esperanto into German, i.e. not a full translation operation.