# MT on-line from CompuServe

*CompuServe, one of the leading on-line service providers, now offers all subscribers a rapid response automatic translation service. The CompuServe Document Translation Service was launched on July 27, 1995, and offers translation in the language pairs English to French, English to German, English to Spanish, French to English, German to English, and Spanish to English.*

The CompuServe information makes it perfectly clear to would-be users that the translation will be imperfect, but is intended essentially to allow the user to understand the meaning of a document: "Use the automated machine translation software to produce a draft translation of your document, which is usually sufficient in quality for someone to understand the document's meaning, though the translation may contain some grammatical errors and mistranslated words. Machine translation is a new technology and may not generate exact translations, but is usually understandable and useful in cross-cultural communications".

The cost of unedited machine translation is three cents a word. Further along in the explanations for users the essential message is punched home again.

"Machine translation is the translation of texts using computers, and is one of the oldest and most challenging areas of artificial intelligence research. Translation by computer is difficult because human language is often complex and ambiguous. Many words and phrases can have different meanings depending on their context. Vocabulary evolves continuously, with new words being added, and others disappearing from common usage. In addition long and complex sentences are difficult for computers to analyse. As a result the quality of translations produced by computers is much lower than human translations. But in many cases, machine translation can produce an understandable rough draft translation in a very short time."

In the opinion of *Language International* you cannot say fairer than that, but that has not stopped the new service coming in for its share of abuse from a small number of professional translators, particularly on the FLEFO bulletin board forum.

## How does it work?

Tests done by Praetorius on the fully automatic translation showed results unsurprising to anyone familiar with machine translation: that the quality was always well below that of a professional translator, that simple texts produced understandable material, while complex texts, particularly in the (notoriously difficult) German to English language pair, ranged from understandable with difficulty to something approaching gibberish.

The user can, however, exercise an option to have post-edited machine translation at ten cents a word (with a $50 minimum charge).

"For documents that require a more accurate translation, a translation editing service is also available for an additional cost. If you select editing, your document will be machine translated, then forwarded to a professional translator who will correct any errors. The resulting translation will be of sufficient quality to be readily understood by a native speaker of the target language.

The post-editing service is in fact provided by Linguistic Systems, Inc. of Cambridge, Massachussetts, a well-established and respected US translation company (founded in 1967).

Examples of translations, both edited and unedited, are given as an option of the menu for the service. There is also an option to provide feedback to the controllers of the service.

The aim is to complete most translations within 48 hours. *Language International*'s tests showed that raw machine translation of short texts had a response time of a few minutes.

When the document is submitted for translation, the CDTS automatically counts the number of words in this source document, calculates a total charge,



*Geoffrey Kingscott interviews Mary Flanagan*

and asks the user to confirm acceptance of this charge (billed to his or her CompuServe account) before translation is commenced.

To access the CompuServe Document Translation Service, use GO TRANSLATE, GO TRADUIRE, GO TRADUCIR, or GO ÜBERSETZEN.

## Preparing a new service

In November, with the service in full-scale operation, *Language International*'s Geoffrey Kingscott visited the headquarters of CDTS at Waltham, Massachussetts, to visit its director, *Dr Mary Flanagan*, who has headed the project since it began in November 1992. Mary Flanagan is herself a graduate in linguistics from Georgetown University. Her official title inside CompuServe is Group leader, Natural Language Technologies.

She explained how she had gone about the business of organising the on-line service. The first task back in 1992, after Mary Flanagan had familiarised herself with the on-line environment, was to look at machine translation systems which were then on the market and which could be adapted for on-line work. The various systems, ten in all, were subjected to a 'black-box' evaluation process, checking out their performance in intelligibility, accuracy and style.

Eventually three systems were short-listed (Systran, Logos and Intergraph) and subjected to a second set of 'glass-box' tests. Whereas a 'black-box' evaluation is only interested in the product which emerges from the system, a 'glass-box' evaluation is interested in how the system works. The accessibility of the dictionaries, for example, was important for CompuServe, and here Intergraph

scored well; the restrictions which Logos and Systran placed on how verbs could be added disadvantaged them in CompuServe's particular kind of application. Systran had performed well on output, but the need for a 3270 emulation board to make a PC perform like a mainframe was unappealing in the universal on-line situation. Other issues concerned the training required for use, and questions of cost and speed. Although Intergraph itself had limitations (particularly concerning the size of its dictionaries) it was judged to be most suitable all-round for the CompuServe application, and licensing negotiations were started in September 1993.

In CompuServe itself there were discussions about where to place the translation service, and while these went on work began on adapting the dictionaries for on-line use. No machine translation system had dictionaries which were designed for the on-line application, a fast-moving world where language is unrestricted — not to say uninhibited. The CompuServe lexicographers (recruited from mother-tongue speakers with linguistics qualifications) quickly found, for example, that there were five different common ways of writing "Thanks".

Text analysis tools were created inside CompuServe to look at discourse frequency levels and one-, two- and three-word patterns.

The translation system was first introduced into Computer support forums, starting with MACSIM, where the discourse was more easy to forecast, for the translation of messages which had been downloaded and subjected to some pre-editing. The pre-editing procedures were developed by Phil Jensen.

An alpha test was carried out for one month in 1994 on CompuServe employees, then a beta test for one month on 100 CompServe members plus some internal users. By August 1994 they were ready to try the system out on MACSIM, which received inquiries in French and German.

The MACSIM experiment was judged a success, with the members of that forum showing a positive response. There were some complaints about the quality of the translation, but the general feeling was that although the translations were "not pretty" they were adequate for the purpose.

In February 1995 a further experiment was launched, a much more testing one, with the introduction of the translation system into the World Community Forum, to facilitate multilingual discussion. This has 35,000 members, and the experiment proved quite successful, some 30,000 words a day being translated. This forum has a lot of members in Europe, particularly in Germany, and German is the second language (after English) in use on the forum.

Mary Flanagan has now identified a cycle of behaviour when the translation system is introduced into a new forum. The first reaction is enthusiastic — "Wow! we can use our own language". Then there is a second phase, of the complaints about the "awfulness" of the translations, especially from more sophisticated users (the French become particularly caustic). In the third phase users start to play around with the software, typing in messages in multiple languages to see what the system can come up with, trying to think of difficult challenges. In the fourth phase everyone settles down to using the translation, coping with the quality problems and treating it as yet one more tool for communication.

**Start and future**

The universal CDTS service began on July 27, 1995 in the six language pairs mentioned. It quickly reached a figure of 250,000 words a month, which is gradually increasing. English is the source language for 56% of the translations, German for 27%, French for 11% and Spanish for 5%. Of the requests received, 89% are for raw translation, and 11% for post-edited. The bigger jobs tend to be the ones that get post-edited, and a pattern of repeat orders is now being seen. CompuServe takes great care to protect the confidentiality of all material sent to it for translation.

Mary Flanagan has identified a number of tasks ahead for her unit. She is far from satisfied with the quality of raw machine translation, and would like to do something about improving the situation in regard to the representation of foreign language diacritics. The unit is developing a language learning product which integrates a number of computer-assisted language learning functions, and May 1996 has been fixed as a (hopeful) target date.

There is also a lot of interest in Japanese as an on-line language, and some interest in Portuguese.

The core group is surprisingly small, consisting of, apart from Dr Flanagan, two software engineers, and two lexicographers (Alexandra Kokovidis for German, and Sophie Toole for French), with a current vacancy for a Spanish lexicographer.

Mary Flanagan has now become philosophical about the English translators who have attacked her on the FLEFO forum; there was even at one time a spate of hate mail, but this now seems to have died down.

"I think on-line MT is here to stay," she says. "We are making a profit. Not many people have found a way to do that in machine translation. There is a great demand on-line for information scanning which can be done quickly. That will continue to grow. There is not much competition out there yet. But that will come. MT technology will improve, in speed and quality."