SYSTRAN on PC imminent

SYSTRAN, the doyen of commercially marketed machine translation systems, is to be launched shortly in a PC Windows version. This major breakthrough could mean the dawn of a new era in machine translation, with much greater accessibility to this powerful system.

Report by Geoffrey Kingscott

Jean Gachot

This is just one of a number of moves to widen the accessibility of SYSTRAN. Last year a version was brought out on IBM PS/2, and a new, and much improved system is now available on the French Minitel network.

As news of these major developments filtered through to the *Language International* office, it was obviously time I made one of my occasional visits to the SYSTRAN headquarters in Soisy-sous-Montmorency, north of Paris, to check them out for our readers. On a hot summer's day in July I was received there by *Jean Gachot*, SYSTRAN's chairman and managing director (his son, *Denis Gachot*, runs the American SYSTRAN operation in La Jolla, California) and by manager *Omar Abdulhak*.

The first point of call was to see the new Minitel system, which by the time this issue of *Language International* is out will have come into operation alongside the existing Minitel connection. The new system offers an astonishing 30 language pairs and can be called up by simply dialling 3617 on the public Minitel system.

One of the things I have always found impressive about the SYSTRAN operation is that they allow the visitor total freedom to experiment with the system, to choose whatever language pair one likes, and to enter for translation any phrase which comes to mind.

Now, on this planet Earth I must rank as one of those who has had more MT systems demonstrated to him than almost any other mortal, but it is not very often I am allowed loose on the keyboard. But, and this cannot be emphasised too strongly to any *Language International* readers who are thinking of going into MT, demonstrations by salesmen of machine translation systems which use texts known to them in advance are worthless; any system can be tweaked to make a good showing on any particular text.

So, it was very courageous of M. Gachot to allow me to play freely with the system, especially as

I am no novice at this game. I know many of the language quirks which cause headaches to machine translation systems — the ambiguous noun and adjective strings in English; the ungrammatical sentences, preferably without a verb; French sentences with confusing uses of *de*; exploiting the polysemy of many common German words, etc. But, whether SYSTRAN was successful or not with whatever I fed in, the results were printed out and handed over to me to analyse at my leisure.

What I found was that it made a brave shot at some of my deliberately difficult phrases, but did not usually solve the problem. However, it did have a very high hit rate for straightforward sentences. Obviously the translations were rather literal, but at least SYSTRAN are honest about this — the idea of the Minitel service is to enable people to understand a foreign text, they say, it is not there to be used as a professional "translator". However, M. Gachot adds mischievously that French professional translators do have the advantage that they can use SYSTRAN via Minitel without anyone knowing!

And sometimes the SYSTRAN translation was not that literal, with some clever switches to target language text word order, and, on one occasion (I was trying it out with that notoriously difficult sentence — *No electric-passenger carrying vehicles are allowed past this point*), it changed the passive voice to an active form with *On* and came up with different rendering of the negative, *On ne permet aucun véhicule passager-portant après ce point*. This was an inelegant but not inaccurate rendering of a complex phrase.

The bane of computers, as of human translators, is the slapdash source text, so I tried it out with that trendy fashion in English, to make nouns out of adjectives: *The prime minister returned from the Corfu summit with his reputation for toughness at a new high* — which came out as *Le premier ministre est retourné du sommet de Corfu avec sa réputation pour la dureté à une nouvelle haute*.

There is a difference in performance between language pairs (which SYSTRAN freely acknowledge), with French to English giving the best results.

A major advantage of SYSTRAN has always been the size and sophistication of its dictionaries.

even though the dictionaries are still constantly being updated in the light of translation performance. "Dictionaries are like Gruyère, they are full of holes", commented M. Gachot.

German has always been a difficult language to translate automatically, partly because of the polysemy mentioned above. The system did not translate correctly a sentence of mine which had *Bestand* in the sense of "stock", but as soon as I substituted *Lagerbestand* it got it quite right. The problem of polysemy is counteracted to some extent by the choice of subject domain fields offered to the user when he connects up to the Minitel service.

Another improvement to the Minitel service which is in the pipeline is a Minitel emulator on PC, so that text for translation can be prepared in advance, and waiting times shortened.

One problem I found was that the system was very unforgiving to typing mistakes (which were frequent, as I struggled with an AZERTY keyboard instead of my accustomed QUERTY). However, Jean Gachot said they were aware of this problem, and are working on attaching a spelling checker to the service which would automatically correct some of the more common or more gross errors.

I mentioned the difficulties we have had in connecting up to Minitel from the UK. It is one of these things that is theoretically possible, so everyone tells us, and perhaps some francophile computer buff somewhere in England has managed it, but the key has so far eluded us at *Praetorius*. M. Gachot said that it is now possible to get through to Minitel by telephone (the number being +33 36 43 15 15).

The next visit was to see the PC version in action. The whole SYSTRAN system has been re-written in C language, a massive operation, to allow PC working. As far as I can judge from the demonstrations, the full Systran system is available. A name for it has not yet been decided — *Il est né, le bébé, mais il n'est pas encore baptisé*, commented Jean Gachot.

The system will work on a 486 Mb machine, and may be on the market any time from September 1994. Experience in the United States SYSTRAN office indicates that translation speed on a PC will be of the order of 100 pages an hour, compared with 400 pages an hour on mainframe. The difference between 100 and 400 pages is of course insignificant, since the bottlenecks will occur in pre-processing and post-processing.

The next exciting development could be to link up with a speech recognition system. Many people think that speech recognition and generation is the next big IT development, with speech-to-text machines, from IBM, Philips and Dragon, about to hit the market in a big way.

SYSTRAN have looked at all three systems, and like the look of the IBM system, even though it is likely to require pauses between detached words in the first instance. SYSTRAN are experimenting with a system whereby one interlocutor speaks his language into a speech-to-text machine, the text is translated by SYSTRAN and transmitted down the telephone lines to the other interlocutor, who receives a spoken language version in his own tongue from a text-to-speech generation unit. In the first stage the experiments will concentrate on the language pairs between English, French, German and Spanish.

Later, over lunch, following up my interview with him of two years ago, I put the big question to M. Gachot, as to how he saw the future of machine translation.

He began by pointing out that every company which had tried to produce a sophisticated MT system, capable of achieving reasonably high levels of quality, had lost money. Everyone of course maintained their own system was the best, but the uncomfortable truth was that the investment required to produce a viable system of sufficient quality to be widely applicable was so enormous that probably no single company would be able to sustain the continued capital expenditure necessary.

Although himself a fervent believer in the free market system, he had come to the conclusion that machine translation was perhaps one field where competition hampered progress, rather than stimulated it. The turnover achieved by the companies selling sophisticated MT systems was ridiculously small compared with the investment they had made, and the investment they needed to continue to make. No-one could hope to make a proper return on investment in ten years, far longer than in almost any other sphere of commercial activity.

The capital investment for a new language pair in an existing system, for example, would be of the order of 8 million French francs, plus another 2 million for marketing. One-third of these costs would be for testing. SYSTRAN's turnover in Europe was something like 2 million francs a year at present, although next year it should rise to between 3 to 4 million. An investment of that order was practically impossible to recover, even over the long-term. The real need was for a world company which could co-ordinate the activitiesof the leading players in this field.

Could such a world company (*société mondiale*) ever get off the ground? Only, he thought, if there was a new willingness on the part of MT developers to co-operate. But perhaps the mood was changing, and it could be that next year's MT Summit in Luxembourg could provide an opportunity to look at the possibilities of collaboration.