

Machine translation in action

METAL at Boehringer Ingelheim

At one of the many conferences I attended in 1995 I found myself listening to a speaker from a German university who was indicating that machine translation (MT) was not yet a usable instrument. When question time came I got ready to intervene, to say that I knew of a number of successful applications, but another member of the audience was already on his feet. Mr Alain Paillet, head of corporate language services at the major German pharmaceutical company Boehringer Ingelheim, told the speaker that not only was his department successfully using machine translation, but he could not imagine them coping without it.



Alain Paillet

agreed, and so on a wintry day early in 1996 I made my way to the small Rhine town of Ingelheim, where a quarter of the town's geographical area is taken up by the extensive Boehringer works.

Incidentally the company is always known as Boehringer Ingelheim to distinguish it from another pharmaceutical company, Boehringer Mannheim (the two companies, which are quite separate, were founded by two brothers).

The porter at the gate, quickly deducing my nationality despite my carefully worded inquiry in German, greeted me in English, and gave me my identification documentation, equally in English. In the main entrance hall to which I was directed large panels set out the text of the company's mission statement and its vision for the future — all in English. I inquired about this phenomenon and ascertained that English has been chosen as Boehringer's corporate language, and is used for all international communications and major company documents as a matter of course.

The language service

By no means does this policy make things easier for

Mr Paillet's forthright defence of MT led me to suggest, when we met up after the session, that an article in *Language International* on the Boehringer translation operation could look at how MT can be a practical tool in a working environment. He readily

agreed, and so on a wintry day early in 1996 I made my way to the small Rhine town of Ingelheim, where a quarter of the town's geographical area is taken up by the extensive Boehringer works.

Incidentally the company is always known as Boehringer Ingelheim to distinguish it from another pharmaceutical company, Boehringer Mannheim (the two companies, which are quite separate, were founded by two brothers).

The porter at the gate, quickly deducing my nationality despite my carefully worded inquiry in German, greeted me in English, and gave me my identification documentation, equally in English. In the main entrance hall to which I was directed large panels set out the text of the company's mission statement and its vision for the future — all in English. I inquired about this phenomenon and ascertained that English has been chosen as Boehringer's corporate language, and is used for all international communications and major company documents as a matter of course.

The porter at the gate, quickly deducing my nationality despite my carefully worded inquiry in German, greeted me in English, and gave me my identification documentation, equally in English. In the main entrance hall to which I was directed large panels set out the text of the company's mission statement and its vision for the future — all in English. I inquired about this phenomenon and ascertained that English has been chosen as Boehringer's corporate language, and is used for all international communications and major company documents as a matter of course.

the corporate language service, since it means their work encompasses many tasks besides translation, and particularly editing work to make sure the English used is always of a professional standard. This is why the department is known as language services rather than as a translation service (*Sprachendienst* rather than *Übersetzungsdienst*).

Mr Paillet was recruited by Boehringer in 1979 to set up the department. Before that time translation work was carried out in the company, but there was no central structure. Mr Paillet, a Frenchman, fluent in German and English, a graduate of the technical translation course of the University of Mainz at Gernersheim, and who had previously been working at the German road traffic research institute (*Bundesanstalt für Strassenwesen*) gathered together the translators already working at Boehringer and set about creating an infrastructure (centralised administration, common terminology, rationalisation of equipment, elimination of duplication of effort, etc.).

The Boehringer company, though still family-owned, is today a huge international conglomerate, present on all five continents, with a total of 23,000 staff. It is one of the top 20 pharmaceutical companies in the world, and the leading over-the-counter (OTC) medicines supplier in Germany. Providing a translation and general language service is therefore a major undertaking, and requires a high degree of efficiency and ability to react to requirements. Translation itself is carried out at three sites, Ingelheim itself (nine language staff), UK (four staff) and Barcelona (four staff), but all belong to the Language Services Department and report to Mr Paillet. "Due to modern communications, such as telephone, e-mail and facsimile, it does not make any difference whether the people you are working with are next door or 2,000 km away", he says, "we all count as one unit".

Geoffrey
Kingscott

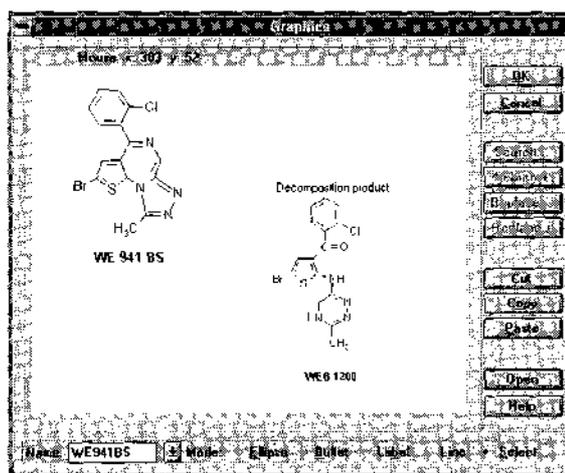
The 15 years Mr Paillet has been in charge have seen a revolution in the way translations are processed. "When we started at the beginning of the 1980s the translators were using typewriters", he recalls. "Terminology was kept on card index cards in drawers. There was little thought of re-use of resources, and because of the previous lack of coordination, some texts had been translated two or three times."

The terminology system

Personal computers were introduced in 1987, and Mr Paillet immediately began to look for a tool beyond the simple word processing which was what was mostly on offer at that time. The concept he was trying to achieve is what we would now call multi-tasking, though it was to be some years before that term came into vogue. A start was made with a tool which could manage terminology, taking an early office management system known as **Deskview** from the American company Quarterdeck, and working to continually refine and improve its effectiveness. Before long the system they had evolved had become indispensable to the working of the department.

When Windows came along at the very end of the 1980s, the department sought to create a Windows version of their termbase. Mr Paillet and his colleagues established a specification as to what a translator would need from such a system, and a computational linguist at the University of Mainz at Gemersheim, Mr V. Srinivasan, then produced a customised system, written in Prolog, with a number of Natural Language Processing (NLP) features. Though it was designed with the five main Boehringer working languages — English, German, French, Italian, Spanish — in mind, provision was also made for other languages and other scripts, including Chinese and Japanese (both Kanji and Katakana). In common with other translation departments throughout Europe, the Boehringer operation has found that the number of language combinations has increased in recent years, and that the rate of increase is accelerating.

The system is installed on all three of the department's sites. Entries are added at any of the sites using a Logfile tool, and transmitted via e-mail. One member of staff is in charge of maintaining the terminology database. There are strict and carefully defined rules for the inputting of new material. Entries for each language allow for source, definition and context information (the context box is of unrestricted length).



The information in the terminology system is now so sophisticated that it has begun to acquire a normative function within the company, and other departments are learning to use it (under Read Only licences). An example which had occurred not long before my visit was a query from the patent department whether there was such a term as *Stant*. Mr Paillet, who at one time knew more or less every term in the database, could not remember this term

being there, but when a search was made there it was, with its definition and foreign language equivalences. A *Stant* is in fact an internal supporting structure to keep blood vessels open (the actual definition in the

machine translation can never replace the human translator, but can only make his or her work more effective

database is much more detailed).

Work continues constantly on improving the database with a view to enhancing its usefulness to translators. Thus work is proceeding on identifying entries through context, on better cross-referencing, on improving and extending the graphics information. Graphics are already in the system, as our illustration shows, and are invaluable to translators (the language of the text on graphics can be changed at will). The graphics in the system (based on **Microsoft Draw**) are also a documentation source in themselves, and can be pasted into a document if required. The system now has over 40,000 entries (the total had reached 43,841 on the day of my visit), each entry being in at least two languages, and many being in the five main Boehringer working languages. Some of the entries may be duplicated, the duplicates lingering from pre-network days, but these duplicates are steadily being weeded out. Abbreviations, the curse of the translator, are also covered, and there are 1,504 of them in the database. Entries are categorised into over 30 sub-domains

(interestingly the term 'drawers', a hangover from card index days, is used). Letter-frequency algorithms for each language speed up the search routine.

Terminology for use by human translators needs to be structured somewhat differently from terminology for use in a machine translation system, and work is now going forward to have an automatic program which will allow the information in one system to be converted automatically for use in the other.

METAL

Which brings us to the **METAL** system, which is regarded by Mr Paillet and his colleagues as simply another tool — albeit a very powerful one — in their armoury of translation aids. They began to evaluate the system in 1990 for German to English and German to Spanish. The purpose of the evaluation was to determine whether a machine translation system could be used effectively for the translation of documents of a technical nature.

The experience of using **METAL** has confirmed the assumptions that were made before it was installed. One is the importance of text typology: certain types of text are far more suited to machine translation than others. The second is that machine translation can never replace the human translator, but can only make his or her work more effective. And the third is that machine translation output must always be post-edited.

But for jobs involving a large amount of documentation, such as drug validation documents, using machine translation undoubtedly speeds up the operation. Between November 11-16, 1995, the department had to translate 200 pages of Standard Operating Procedures for information purposes, and the **METAL** system, under its chief operator, *Yvonne McNeish*, handled the work triumphantly. Compared to the potential cost to the Boehringer company which a delay in processing that documentation would have caused, the actual cost of the operation was trifling.

Before the system was integrated into the department's operation it was tested for six months. It was tested in different conditions, and given some difficult tasks to perform, in order to ascertain just what it could do and what was beyond its powers.

One of the reasons Mr Paillet sees for the comparative slow advance of machine translation in industry is the slowness of developers in making their systems more user-friendly, in the sense of designing them so that they can be integrated into existing

customer environments, electronic data processing packages, terminology databases etc. And vendors are only now beginning to realise that it was not enough to offer an MT system equipped only with a basic vocabulary; domain-specific and customer-specific terminology is essential if the system is to work. There are terminology resources available if the vendors had only taken the trouble to look. Why should it always be the customer who was responsible for getting specialised terminology into the system? Boehringer have a name for programs which have not been geared to the customer — banana software. They feel the vendor is selling them green bananas, and the customer has to see to the ripening before they can be used.

*the vendor is selling green bananas, and
the customer has to see to the ripening
before they can be used*

The other major grievance which Mr Paillet has is the public ignorance about machine translation. Either people have naive expectations, thinking it is a magic machine which can perform the whole

translation process in a twinkling, or that it is a chimera which will never produce results. About a week before my visit there had appeared an astonishingly sweeping dismissal of machine translation by *Mr Jürgen Kern*, president of the German translators' association (BDÜ), which had appeared in the BDÜ journal *Mitteilungsblatt*.

In fact Boehringer's experience confirms the experience of other users I have met, that the more one uses MT the more effective it becomes. The best systems are therefore those which allow the user to "teach" the system. **METAL** is one of those which allows the user to code new lexis, and this coding is of paramount importance. The Boehringer experience is that of the mistakes made by the system in actual translation, 50% of them could be attributed to wrong or inadequate coding. An incorrect verb-frame coding would lead to erroneous analysis, and therefore erroneous generation in the target language. The **METAL** system has some 500 grammar rules, which are applied to the analysis of sentence pattern until one matches; parsing is then carried out, the sentence analysed working outwards from word to complete sentence, all of this in split seconds.

The coding process usually involves verbs, nouns and adjectives; other parts of speech are usually already coded in the basic system. Experience shows that while there are always problems in machine translation, these problems become manageable as soon as the system has been 'taught' to operate in a 'sub-language' (*Fachsprache*) where the terminology has specific meaning. The text types which work best in machine translation are those of a declarative

character, such as instructions for use, and where there is a relatively straightforward sentence structure.

Average output from the system is around 30 pages an hour, which then go for post-editing. About ten per cent of the department's workload goes through **METAL**, though there can be periods of very intensive use. "It's nothing more, and nothing less, than a tool", comments Mr Paillet.

The future

So satisfied are Boehringer with the results of their machine translation that the department, which is already a separate profit centre within the company, is examining whether to offer the service generally to other pharmaceutical companies around the world, through some sort of e-mail access.

Mr Paillet sees the future as bringing about an increase in the natural processing, using a variety of tools, some of which are still to come on the market. Communication was increasing all the time, and departments such as his had to find ways of coping with the demand. He believes that the translation

profession must do more to enhance customer awareness of its work.

This has certainly been done at Boehringer. The corporate language services department has consistently maintained a high profile, and has enthusiastically embraced the new recently enunciated company philosophy, which involves concepts such as vision, leadership and value through innovation. On November 8, 1995 there was a company-wide innovation day, with every department making a presentation. The language services' presentation, which showed **METAL**, the **termbase**, the **Compare and translation** memory systems, together with two light-hearted quizzes, was highly successful, and was busy all day. Similar exhibitions were held on all three sites.

As a result of this and another initiatives, the language services department is regarded as something of a showpiece at Boehringer, and provides an object lesson of how the translation profession, by embracing the challenge of innovation while maintaining standards, can raise its profile.