

5

Computer-assisted translation at ITT

A talk given at the British Computer Society
Natural Language Translation Specialist Group
15th April 1982

Ulla Magnusson-Murray

Introduction

Maybe you have come tonight expecting to see, one after the other, samples of high-quality machine-translated output and to hear an account of smooth-running hardware, sophisticated software and ever-smiling translators; not so: I shall talk about problems, because that is what makes life interesting.

I shall discuss the problems facing the human translator, problems facing an in-company translating department, problems with co-ordination, terminology and incomprehensible syntax and - not to forget - problems with impossible dead-lines.

But before we get totally submerged in gloom and despair, let's try to look at a solution to the eternal language problem; there would be no need for translators to "slog" away at 3000 words a day or, for that matter, to invest in computers if we were to adopt Allan Coren's methodology as exemplified in Figure 1.

Background

In the absence of a fully-developed Franglais, Swenglish or what have you, what made ITT decide to consider the idea of computer-assisted translation? To answer that question, I would first of all have to describe the following:

- a) the function of the ITT Technical Translations Centre within the framework of ITT
- b) the problems which prompted changes in our approach to translations and other language-related areas.

Our Centre was established over 5 years ago to provide a technical service to ITT companies all over Europe. As part of the ITT Europe Engineering Support Centre, we carry out translations of research papers, bids and tenders, technical specifications and technical product descriptions, but mainly of large product and customer documentation for major systems, such as System 12, which is ITT's recently developed digital telecommunications system. There are naturally many difficulties involved in translating such vast quantities of documentation into, e.g. Spanish, Norwegian or Italian - let alone Arabic and Russian. For instance, translators with the right combination of technical expertise and literacy must be available, glossaries must be compiled, queries sorted out, and, more often than not, customers do not give due consideration to the time it takes for translating. Our department has acquired a great deal of expertise in translating large system documentation from many ITT companies, and so far, our own expert translators, supplemented by editors, have been

LET'S PARLER FRANÇAIS

Dans le Railway Station

Monsieur: Bonjour.
 Caisser: Bonjour, Monsieur.
 M: Je veux aller a Shrewsbury.
 C: 1er, 2ème, un-way, deux-way, cheap-jour, football-terror spécial...?
 M: Un-way ordinaire, svp.
 c: £19.50.
 M: Pur quartre.
 C: £78.
 M: Ah, mai j'ai un railcard!
 C: Oh, Jesus.
 M: Pour ma fille, j'ai un Teenage-problem railcard.
 C: Un Teenage-Problem Railcard? Vous prenez le Michael?
 M: Mai non. C'est un nouveau railcard. Il garantit que votre adolescent pimplé peut voyager libre a condition que:
 1) il ne demolish pas les fittings
 2) il ne terrorise pas les passagers avec les idees trotskyistes
 3) il ne become pas plaque avec vino Plonquo
 C: Bon. Un ticket free. Et le troisième?
 M: C'est Ingrid, notre au-pair de Stockholm. Pour elle j'ai une Au-Pair Runaround Free-Flirt Railcard. Regardez, sa photo.
 c: Mais c'est une photo de Jimmy Saville!
 M: Toutes les jeunes filles de Suède ressemblent a Jimmy Saville.
 C: L'au-pair Railcard n'existe pas!
 M: En ce cas, je desire un ticket communal pour mois, ma wife et Ingrid. Regrdez, mon Menage-a-Trois Fun-Threesome Railcard. J'ai une photo très intéressante!
 C: OK, OK, vous etes le winner. £29.50.
 M: £29.50? C'est beaucoup.
 C: During notre conversation, on a announced une augmentation du price du 15%. Tough chance. Et si vous ne regardez sharp, on va fermer la station de Shrewsbury. Next svp!

Figure 1

used. Although this is common practice, it has many disadvantages: the time pressure very often compels translators to work up to 12-14 hours a day, and considering translation is a creative art, requiring high mental activity, this pace cannot be kept up efficiently for long periods. If many translators have to be used, the terminology is inconsistent, and there is seldom time to compile product-specific glossaries; editing and verification of translations is therefore necessary, but, since it constitutes a bottleneck, is seldom done efficiently. The result is dis-satisfaction all around and the customer suffers, since his documentation may arrive late and may

require further editing for clarification purposes.

In spite of our own manually compiled product-specific vocabularies, well organised co-ordination and administration, we were still left with two major difficulties:

- a) terminology inconsistency
- b) time scheduling.

The concept of computer-assisted translation appeared to offer a solution to these two problems, and some 2 years ago we therefore decided to look at the market of available translation systems. Being very much at a research and development stage, automatic translation had and has yet to come of age. An accurate evaluation of performance, efficiency, quality of output and cost is difficult and requirements vary from user to user. However, the widespread use of word-processors, phototypesetting equipment, electronic publishing systems and international transmission links, etc., has removed many of the limitations for applications of computer-assisted translations.

System Evaluation

On the basis of our own translating experience, we chose the following criteria in our search for a suitable system:

1. Having realised that top-quality translated output was unrealistic, we needed a system that would produce 70-80% accuracy (after vocabulary entry) and the remainder would be left to the expert translator.
2. The system should have the capacity to store a large number of words, which could be easily updated or deleted by us.
3. The system should allow us to enter our own ITT product-specific glossaries.
4. It should offer a substantially improved turn-around time.
5. The interaction with the machine should be "user-friendly", i.e. via natural language, and the equipment should be ergonomically designed. The various stages involved in operating the system should be stimulating, and require the expertise of a translator. In other words, it was, or is not, my intention to do translators out of a job; as a matter of fact, I wish to make a plea for the status of the translator in this country: as long as employers equate a translator to an office clerk, there can never be job satisfaction. Introducing computers as a tool for the translator should, therefore, not only make his task more interesting and forward-looking, but offer him the professional recognition he deserves.
6. The system should allow us to update the programs in order to improve the quality of the translated output, and also offer interesting linguistic work for the translator.
7. The software should be compatible with the computers that already exist in our building at ITTE-ESC.
8. The system should be capable of linking to existing ITT Computer networks in the UK and on the Continent.
9. The system should offer substantial long-term cost-savings.

As if this is not enough, generally speaking, there are many more

features that one could wish a translation system to perform, but these are not part of this talk.

The Weidner System

The system we eventually chose to run our evaluation programmes on was the Weidner system. It meets, in theory, most of our criteria, with some exceptions. For instance, we were immediately told by the supplier that ITT could not update the software programs themselves.

The other criteria are now being thoroughly investigated in practice by us. At this point, it may be appropriate to clarify a misunderstanding: the magazine "Computing" reported sometime ago that ITT has spent a considerable amount of money to buy the Weidner system. This is not strictly true; in fact we are renting it and I will now explain what type of installation we have.

The ITT Installation

It was originally the intention to use the translation programmes on our existing VAX-11/780 computer. However, the programs, being developed for the PDP series, would require adaption to suit the VAX. This is now being done in the US at Weidner and by September this year, we hope to use the VAX. For our operational evaluation programme, we are now using a PDP 11/34 and our configuration is as in Figure 2 (q.v.).

CONFIGURATION

- PDP 11/34 and console (VAX by September)
- 2 disc drives
- 2 terminals
- 1 multi-lingual printer
- English/French/English
English/German/English Software Packages
English/Spanish/English
- A number of translators with square eyes, enthusiasm and
patience like saints

Figure 2

Without the last ingredient, I probably would not have had a system today. The translators' keen interest in the project has kept them standing by, while a string of hardware engineers have been busy sorting out our teething problems.

In February we had a 2-week training period on the German program and

we have just completed the Spanish training. The last 3 weeks have been spent on general familiarisation, inputting of texts, finding out about the structure of the vocabulary in the various directions, inputting ITT vocabulary and giving umpteen demonstrations.

Operational Evaluation

Before discussing samples of translations we have done so far, I would like to describe some of the work involved in evaluating the system as being suitable for use within ITT. The vast task of organising the evaluation programme is now in preparation and we shall concern ourselves with tasks such as those in Figure 3 (q.v.).

Operational Evaluation of the ITT CAT System

- How long is the period of familiarisation with the programs?
- What is the intellegibility of the translated text?
- what is the intellegibility of the source text?
- How much knowledge of computers and peripherals is necessary?
- What is the time taken for:
 - keying-in of, say, 1000 words (technical text)
 - vocabulary search for same text
 - dictionary entry of words not found in same text
 - processing time
 - post-editing time?
- what type of errors is most common?
- What type of errors/difficulties can we deal with ourselves (percentage)?
- What type of errors/difficulties can only be dealt with by supplier (percentage)?
- Having decided that technical text is well suited to CAT, can the source text be improved by using "Controlled Language" (i.e. pre-editing)?
- Are customers interested in raw output to edit themselves?
- What is break-even point at which CAT becomes economical for ITT?

Figure 3

These are only some the aspects we shall be looking into and, reallistically speaking, we do not expect to have any calculable results for another 3-6 months.

However, since most of you have come to see results, I shall now show some typical examples of texts that we, as a department, work with. The texts are Spanish into English. Figure 4 shows you the original input, Figure 5 the raw output (untouched by human hand), whereas Figure 6 is the result after being edited by my translators.

Original Spanish Input

1. Preambulo

El sistema 1240 es un sistema digital con control distribuido en el que la red de conexión es el soporte de la comunicación entre los microprocesadores que controlan una llamada. Consecuentemente, la red de conexión ha sido diseñada como una red "virtualmente sin bloqueo" para asegurar una perfecta comunicación entre los elementos de la unidad de control

2. Tipos de conexiones y método de búsqueda de camino

La red de conexión es el soporte de la comunicación entre líneas, enlaces y circuitos de servicio, y entre los microprocesadores.

El método de búsqueda de camino es paso a paso con posibilidad de 4 intentos en las conexiones de conversación y señalización y de 16 intentos en las conexiones de control. El procedimiento para establecer el primer intento es como sigue: (ver Fig. 1).

Dependiendo del tamaño de la central y de la posición relativa de los dos interfaces terminales, el camino se refleja en la matriz de acceso o en la primera, segunda o tercera etapa de la Unidad de Grupo.

Figure 4

Judging by these texts, the 70-80% accuracy has not yet been attained, but one must keep in mind that very little of ITT's technical vocabulary has been entered so far.

As I mentioned before, 3-6 months of evaluation are needed to give a full picture, but some language-specific problems have already emerged: for instance, we are having difficulties with the imperative form from Spanish into English. This verb form occurs frequently in our technical texts, especially in installation and maintenance manuals and we are therefore working on a solution at the moment. The program also has difficulties with the fact that Spanish very seldom uses personal

Raw English Translation

1. Preamble

The system 1240 is a digital system with distributed control in which the switching network is the support of the communication between the microprocessors that controls a call. Therefore, the switching network has been designed like a network "virtually block-free" to ensure a perfect communication between the elements of the control unit.

2. Types of switching and path finding method.

The switching network is the support of the communication between lines, trunks and service circuits, and between the microprocessors.

The path finding method is step-by-step with possibility of 4 attempts in the calling and signalling switches and of 16 attempts in the control switches. The procedure to establish the first attempt is as it follows: (to see Fig. 1).

Depending on the size of the exchange and of the relative position of the two terminal interfaces, the path it is reflected in the access matrix or in the first, second or third stage of the Group Unit.

Figure 5

pronouns together with a verb. In German into English we find that verbs very often have to have a separate entry in addition to the infinitive when it is used adjectivally, e.g. Ein bestimmter Amt=A particular call. The German passive form is also causing some headaches. In general we find that questions about grammatical properties that the translator must answer are not specific enough and could certainly be improved. But amidst this, there have been moments of great excitement: there is no controlling us when sentences are produced that require virtually no editing whatsoever!

Commercial Applications

If we are to keep the system after the evaluation period, we must consider the commercial application: in our case technical documentation.

Our yearly turn-over before the computer installation was around 600,000-800,000 words. This figure includes all languages we deal with. To make our system cost-effective at today's figures, we need to translate approximately 1.5-2 million words a year in total. However, the initial cost of acquiring the programs will decrease and so will the cost for CPU time generally. With these turn-over figures in mind and the expertise we are building up, we are naturally looking for

Edited English Translation

1. Preamble

System 1240 is a digital system with distributed control in which the switching network is the medium of communication between the microprocessors controlling a call. Therefore, the switching network has been designed as a "virtually block-free" network to ensure perfect communication between the elements of the control unit.

2. Types of switching and path finding method.

The switching network is the medium of communication between lines, trunks and service circuits, and between the microprocessors.

The path finding method is step-by-step with the possibility of 4 attempts on the calling and signalling switches and 16 attempts on the control switches. The procedure for establishing the first attempt is as follows (see Fig. 1).

Depending on the size of the exchange and on the relative position of the two terminal interfaces, the path is reflected in the access matrix or in the first, second or third stage of the Group Unit.

Figure 6

large contracts and more customers, inside as well as outside ITT. So far, there has been nothing but keen interest in our work; some scepticism, yes, but mostly from people who imagine that a computer should translate at 100% accuracy and then find that impossible to believe. When the more modest aims are explained, even the most sceptical become converted. Computer-assisted translation to the customer means:

- improved quality through product-specific terminology consistency
- greatly improved turn-around time
- increased export possibilities
- possibility of customer's own editing of raw translated output.

To copy type source text into the machine is neither attractive to the translator nor is it economical. We must therefore make inputting more cost-effective and this can be achieved by interfaces with wordprocessors and typesetting equipment, OCR's and - of course - data links. We are establishing a Wang interface at the moment, inhouse as well as with other ITT companies. When we transfer the translation programs onto the VAX computer, we will automatically have access to the DECNET and the IBM Network via data link. Documents can then be transmitted along a telephone line, translated on our premises and the translated text, in raw or edited form, be transmitted back to the customer. An exciting future for many - others may find the prospects

inhuman and frightening.

Final Remarks

Leaving the results of ITT's evaluation programs aside for a moment, it is, I am sure, apparent that computer-assisted translation is the solution to produce clear, consistent and unambiguous translation of vast technical documentation rapidly and cost effectively. Being out of reach for the small translation agency, from organisational as well as financial reasons, it is ideally suited to large multi-national corporations. As I mentioned earlier, however, without the expertise of highly competent translators, even a multi-national organisation will not do justice to such expensive equipment. How do translators take to these new gadgets? I can only speak for my staff and myself and it being early days, one should not pass comments prematurely. Apart from being an experienced technical translator, the type of person you need for computer-assisted translation must be friendly disposed to computers, regard them as a tool and not as a threat, enjoy a challenge of controlling the system and be prepared to teach it and use it to his own advantage. So far comments from my translators range from:

- the chap who developed this, didn't know the difference between declensions and conjugations!

to:

- Imagine an artsperson like me being able to work with computers!

and

- Can you imagine sitting in front of one of these for the rest of your life!

For me, personally, the excitement does not only rest with the translation program itself, but also provides the challenge of learning about the actual operating system and computers in general.

You may remember that I talked about how we can reach the customer via data links; one can think of a similar arrangement with a free-lance translator. Of course, a translator needs to have a Hewlett Packard terminal at his house and a modem link. He can then use our computer from his home and perform all the tasks the staff translator performs in the office.

After all, as a tool to use in technical translation process, the computer seems to be here to stay - it is up to the translator to employ it in a creative and intellectual way.