[*Terminologie et Traduction*, no.1, 1986]

SYSTRAN - THE TRANSLATOR'S VIEWPOINT

Andrew D.C. Evans
English Translation Division
Commission of the European Communities, Luxembourg

Shortly after I became a translator here at the Commission I was reminded by a colleague a great deal older and wiser than myself that inside every translator, there is a writer trying to get out. As this is scarcely the occasion to embark upon a literary career I shall try to keep to the point.

Commission Systran, as my colleague Ian Pigott told you a few minutes ago, now boasts four operational language pairs, with a further five under development. It is frankly too early for the English-German system to be evaluated: for technical reasons unconnected with Systran, not enough operational output has yet been produced to form a judgment. Nor can I in all conscience claim to represent the views of the French or Italian divisions; I am taking quite enough of a risk in speaking for my British, Irish and Commonwealth colleagues, many of whom are sitting in the gallery waiting to protest if I misrepresent their views. So I speak exclusively of English target language Systran, which, as you realise, means exclusively French source language.

French-English Systran has been available in our translation service for some five years. It made a few friends in its early days, though perhaps fewer real friends than was widely believed: most of my colleagues who cultivated Systran did so for its undeniable entertainment value. Once Systran stopped being fun its acquaintances silently vanished into the mist. A few of us have nevertheless stuck with it, and development work over the past five years has enabled us to reach some fairly firm conclusions about what uses we can find for Systran.

Let me start by the bad news. There are a number of areas where Systran is useless and sometimes positively dangerous. The proceedings of the European Parliament, for example: I mean no criticism of the Members, or of Systran: the speeches are remarkable and sometimes admirable rhetoric, but for the moment their style is beyond Systran.

Oratory is out because it is at the warm end of the emotional scale; at the other end is the icy chill of legal texts - anything of an unequivocal nature. As Sir Bruce Fraser said, "... By normal standards of good writing legal drafting is usually both cumbrous and uncouth... "([1]).

Style is the start (but not the end) of Systran's problems. What we are left with is the middle-of-the-road, everyday output of bureaucracy the world over: administrative reports, statistics, technical reports, minutes, notices of meetings. Hard, cold facts: but not too hard, please, and not too cold.

It was at the point when initial enthusiasm for Systran was starting to wane that our division hit upon the idea of rapid post-editing. Our experiences at that point had not, frankly, been all that positive: although we were willing to go along with Systran development, output could seldom be post-edited to an acceptable standard within the time a competent translator would have taken to do the job from scratch. A solution had to be found which satisfied Systran's hunger for output, and consumed no more resources - that is, translators' time

and patience - than conventional translation. The answer was found in rapid post-editing. It is a system by which everybody gains something; Systran is used and gets feedback for development; we get MT which is an aid to translation, and the requesters derive benefits which will be described by Mr D'Erman and Dr Walker later today and tomorrow.

Rapid post-editing could never have been implemented without high quality word-processing. It quickly became clear in 1980 when the first hard-copy production MT rolled out of the printer that pen-and-paper post-editing was a tedious chore. Due credit must be given to D6 XIII who in 1981 bought a user-friendly but highly sophisticated word-processor suitable for use in post-editing (and, incidentally, in translating), and who have since then vigorously defended their choice, enabling us to post-edit machine translation in the best way we know of.

If I may be allowed a short digression on word-processing for translators and post-editors, I might diffidently suggest that our needs and skills are not identical to those of a commercial office, and that the requisite system must be powerful but simple enough to be used by a non-expert: translators tend to be expert in unexpected areas, but rarely in office skills, computer literacy or (save as a relaxation) cerebral gymnastics.

In fact the word-processing exercises most frequently needed in post-editing are, three. The example of a post-edited text will give you an idea of their share in the total effort needed to make a text comprehensible.

1 Search and global replace. These are boxed. Five examples have occurred in the space of three paragraphs: an acronym with which Systran was unfamiliar, a mistranslation - rang as "row" rather than "rank" or, in this case "status", one economics term, one horror - Asie of the south-is, and one chestnut: the French points de suspension "..." not being removed when it is translated as "etc.;." Terminology or target language construction errors are usually best dealt with comprehensively, making subsequent minor corrections occasionally rather than major ones systematically whilst working through the text.

2 Change word order B-A to A-B, B-C-A to A-B-C, C-A-B to A-B-C, etc. This is particularly because of Systran's difficulties disentangling some of the wilder French concatenations of, for example, nouns and adjectives, and you will see that the routines have been used six times in three paragraphs.

3 Delete. These are shown ruled through, but the boundary between deletion and other forms of text change is uncertain, so all other changes have been marked in the same way. You will note too that notwithstanding the repeated lists, the English is significantly shorter than the raw MT - about 7% in this case. But French is so wordy compared with English that often up to 20% of the bulk of the text can be pruned without any loss in intelligibility, and usually with some gain.

Of the four word-processing systems tested and used by us only the Wang system provided through the Systran team can perform all these functions economically (i.e. quickly and using a minimum of easily-memorable keystrokes).

We have found that rapid post-editing without the word-processors is not a success (although in fairness I must add that our French colleagues do not agree with us). A few of my colleagues use pen and printout, but most prefer to work on screen. For a number of them rapid post-editing has presented adaptation problems of a different kind: one might almost call them ethical. I come from a background where anything lawful which got the work done was fair game - the British Inland Revenue, actually - and since I was fairly new when Systran was introduced I had no trouble lowering my standards to the point at which I could say "half an hour a page, no more, and too bad if any mistakes are left afterwards". Not all my colleagues have this flexibility, and I admire them for it.

An interesting point here is that flexible standards do not mean no standards at all; in fact the evidence suggests that the better (and perhaps faster) a translator is, the better he will be as a post-editor. Dexterity with the word-processor is also critical, and in order to improve such skills we are all encouraged to use the terminals for conventional translation, although Systran takes priority. A good deal of commonsense is the third requirement; and next comes, not quite decisiveness, but confidence: the ability to decide quickly what, if anything, needs to be done to correct an apparent mistake: it's all part of being a competent translator, I suppose. There's no point in dithering trying to decide whether to correct something or not: if it's good enough to leave you in two minds it will probably do; and time spent wavering can be more profitably used to sort out something more deserving. Last, speed: you're working against the clock.

These are the definable qualities: good translation skills; dexterity with the word-processor; commonsense; decisiveness; an ability to work quickly; and I would add a sixth: the nose for a wrong 'un. In many jobs you develop what the French call a déformation professionelle, and in post-editing it is the instinct for a quite innocuous-looking translation where Systran has got hold of the wrong end of one stick, and extrapolated from it a grammatically perfect, stylistically admirable translation which has only one defect: it is totally, completely, utterly wrong. A short example from the same text might be the Importations communautaires des PVD, which is ambiguous enough out of context, but set Systran onto the Community's imports from developing countries, rather than the reverse.

So, with the right translator using the appropriate (i.e. flexible) approach to the right policy (rapid post-editing), and the right word-processors, Systran can be made to work.

We have set a limit of 30 minutes per page, and a target of 20, and within that limit the translator is free to post-edit the text as he thinks fit. The translation is sent out without further revision, and the requester accepts it on the explicit understanding that it does not come with the usual guaranteed accuracy or elegance, although the word-processed presentation does perhaps to some extent compensate for the lower intrinsic quality of the translation. Mr D'Erman will tell you more.

We are some 40 in the English Translation Division, of whom 16 are at present on the list of Systran volunteers. If the word-processor terminals were fully used we could handle 150 pages of rapid post-editing each week. In fact during 1985 we did the equivalent of barely a month's rapid post-editing: 27 documents totalling 605 pages (including a single document I did accounting for a full third of that total); for lack of opportunity four volunteers did no Systran at all; five did only one job, and none did more than four.

Why, then, is it used so little? The first reason is that not all documents we translate start life as French. Only about 35% of them, in fact; and not all of them are suitable, as I have already suggested. The second is that rapid post-editing has not attracted the requesters as had been hoped, a matter on which I am not qualified to comment save to say that the markets for machine translation and human translation are not necessarily one and the same, and that my colleagues and I would regret to see them confused. However, I am certain that there is a market for machine translation within the Commission, and that Mr D'Erman will be able to persuade potential requesters of its advantages to the end-user.

And the third reason for Systran's lack of turnover is that with the best will in the world, a silk purse cannot always be made from a sow's ear. I hasten to say that Systran itself is not necessarily at fault, but there are a number of weaknesses in the production process which diminish Systran's ability to do a good job.

One critical point is the input of the source text. Sometimes it comes straight off a word-processor. Most Commission departments now have word-processing resources, and many texts for translation already exist in machine-readable form. A diskette economy is slowly developing, and Systran is able to benefit from it. The system does have its limitations, however, one of which is the intransigence of the French who seem bent on using accents in their language. Inability to recognize an accent is one of the problems which occurs occasionally in diskette conversion, and usually in optical character recognition. The effects on a machine translation of a system incapable of recognizing the difference between "marche" and "marché" are considerable. In the same order of problems we find the more human errors of input to which we are all susceptible, though perhaps more here than elsewhere, since only in an institution like the Commission is one likely to find a Greek typist keying minutes drafted in French by a Dutchman, recording the comments made in English by a Dane on a joint report by an Italian and a German.

Although there can be problems with the input source text when it is already in machine-readable form, there are even more when it is not. Even who types it is a problem. We have found an answer in the case of one requester, who has been provided with a Wang terminal, but not all requesters have the demand to warrant such an outlay. Our ladies in the English typing pool are having enough trouble keeping up with typing finished English translations, and were certainly not recruited to type French anyway. In fact the French typists usually do the input for us, quid pro quo our pool key in the English text for MT into French. It's not a very satisfactory arrangement, not least because a job in which the typist has no direct interest is not going to get the very best treatment. Occasional carelessness, the odd mis-spelling, an unconventional piece of punctuation, and the quality of the machine translation slides off the bottom of the graph.

Whether we are talking about human translators or computers trying to do the same job, the essential problems of the task are two: the first is that of understanding what the source text is about; and the second that of rendering the ideas it contains into the target language. Our retired head of service here, Albert Bachrach, summed it up with characteristic translator's neatness and precision, and admirable Dutch economy: translation is the art of getting an idea from one language into another. However, ideas are our exclusive preserve as humans: Systran, in the last analysis, is no more intelligent than an abacus. Frequently it doesn't matter, but occasionally one could wish it had some capacity to eliminate the absurd. President Reagan, Governor Reagan, even Mr Reagan, but certainly not Chairman Reagan.

I mentioned the French language itself as a potential cause of trouble, and I should like to spend the last of my allotted 20 minutes on this question. Here is an example of what I mean. It's a letter from the French tax authorities, and in some thousand words it tells me that I haven't yet sent in last year's income tax return, and I had better now do so or else. As I said, for ten years or so I was myself on the staff of the British Revenue, and this is a situation which is not unfamiliar to me. We had the same thing, and my recollection is that Form 155 runs rather more simply:

> "Dear Sir or Madam
> The form for return of income and gains sent to you recently has not yet been received in this office. Will you please now complete, sign, and return it to me. If you have already sent a tax return to another office, please let me know which one.
> Yours faithfully
> HM Inspector of Taxes"

The difference in what might be called reader-friendliness between the two styles makes a big difference to the quality of the Systran output.

French has not undergone the spring-cleaning given to English administrative writing in the last 30 years at the hands of men such as Fowler ([2]), Herbert ([3]) and Gowers ([4]).

In 1946 no less an author than George Orwell laid down six elementary rules of writing ([5]), and Orwell would, I am sure, not mind being parodied in order to sum up Napoleonic administrative draftsmanship:

1 Do not use a metaphor, simile or other figure of speech unless it is already worn out;
2 never use a short word when a long one will do the job worse;
3 never use one word to say something you can expand to a relative clause;
4 never use the active when you can use the passive;
5 use foreign, learned, obscure and jargon words wherever possible;
6 write badly sooner than break any of these rules.

I would not go so far as to say that my Napoleonic letter from Thionville tax office obeyed all these rules: if it did, it would be totally incomprehensible.

Neither do most of the jobs we get for Systran. But it still doesn't really surprise me that Systran was unable to provide a sound translation of a letter which must have been unintelligible to most of the Frenchmen who received it, and I will not try your patience by reading the MT to you.

In answer I have one suggestion to offer. A point now beginning to emerge is that Systran is sometimes too accurate. One of the regular parts of a modern translation course is an essay on the subject "In translation it is better to be approximately right than precisely wrong - discuss". I leave you with the thought that one of the logical consequences of rapid post-editing as a policy is a shift away from the ideal of precise accuracy and its concomitant risk of inaccuracy, towards a vaguer rightness. For ten years now efforts have been directed at getting Systran to provide the most accurate translation possible. As the most successful of Systran applications, rapid post-editing seems to indicate that some change is called for in that policy, and I shall be interested to see in a couple of years whether this is not in fact the right path.

References:

(1) The Complete Plain Words (Second edition, rev. Fraser) HMSO 1973
(2) Modern English Usage (Second Edition, rev. Gowers) OUP 1965, etc.
(3) What a word! Methuen 1949 (any many other works on the subject)
(4) The Complete Plain Words (First edition) HMSO 1954
(5) Politics and the English language, 1946

RAW OUTPUT	RAPID POST-EDITED (20 minutes maximum)
The border between countries industrial advanced and NPI, or NPI and other developing countries, evolve permanently. The countries form, according to the expression by OECD, a "dynamic continuum" in the process of development. They reach for certain products the competitive country row on the world market, but can lose or strengthen this position within the variable time preceding to the evolution of their national productive structures, of the forms of competition on the world market, of the technological transformations, etc. ...	The borderline between advanced industrial countries and NICs, or NICs and other developing countries, changes constantly. The countries form, to use the expression of the OECD, a "dynamic continuum" in the development process. For certain products they attain competitive country status on the world market, but can lose or strengthen this position in time, depending on trends in their national production structures, the forms of competition on the world market, technological change, etc.
In the Sixties, a first group of developing countries emerge on the basis of the growth rates of their internal production and their very high exports. This group comprises four states from Asie of the south - is: Hong Kong, South Korea, Taiwan, Singapore; four countries from Europe of the south: Greece, Yugoslavia, Spain, Portugal; and two from Latin America: Brazil and Mexico.	In the Sixties, a first group of developing countries emerged on the basis of the growth rates of their domestic production and their very high exports. This group comprises four states of South-East Asia: Hong Kong, South Korea, Taiwan and Singapore; four countries from southern Europe: Greece, Yugoslavia, Spain and Portugal; and two from Latin America: Brazil and Mexico.
A second wave of exporting developing countries of manufactured products finished in the decade seventy: the Philippines, Malaysia, Thailand, Macao, Indonesia, India, Argentina, Tunisia, Morocco, Mauritius, etc. ... Some among to soon revolving plated between first NPI and the developed economy markets. First NPI has there, indeed, delocalized or sub-contracted a number of activities under multinationals of the Third World. This strategy facilitated the development of the south-south trade, but it also strengthened the capacity of export of a growing number of developing countries which become competitive countries on the world market.	A second wave of developing countries exporting manufactured goods arose in the seventies: the Philippines, Malaysia, Thailand, Macao, Indonesia, India, Tunisia, Argentina, Morocco, Mauritius, etc. Some appeared as centres trading between the first NICs and the developed-economy markets. Indeed, the first NICs set up or sub-contracted locally a number of activities through Third World multi-nationals. This strategy facilitated the development of south-south trade, but it also strengthened the export capacity of a growing number of developing countries which have become competitive on the world market.