

MACHINE TRANSLATION AND PEOPLE

Veronica Lawson

30 Half Moon Lane
London SE24 9HU
England

This conference, in concentrating on the experience of people, particularly translators, who work with "practical" MT systems, reflects changes over the last year, partly in the availability of MT, but more fundamentally in the relationship between translators and MT: attitudes of translators and to translators. These changes are attributed to wider recognition of the translator's role in providing linguistic feedback (as, for example, in the author's work on MT of patents).

What is machine translation, and what may it be expected to do? The various approaches - from restricted language to the "try anything" system - are introduced, and the different modes (pre-edited, post-edited, interactive and raw) are outlined. MT's place in the translation market is suggested, as are some likely implications, positive and negative, for the people involved.

This paper had best begin with an apologia for its rather prosaic title, "Machine Translation and People". The truth is that by the time the conference steering committee had finished, it was not so much a paper as a patchwork quilt: the background to the conference, a child's guide to machine translation, possible effects on the market and implications for the translation profession. If there is a common thread, it lies in the relationship between people and machine translation (MT). In the same way this conference itself - "Practical Experience of Machine Translation" - concentrates primarily not on theory or even hardware, but on the experience of the people, particularly translators, who work with "practical" MT systems (production systems in regular practical use, as opposed to systems used for research). Practically all of our speakers have worked with MT. In this emphasis our programme reflects changes over the last year, partly in the availability of MT, but more fundamentally in the relationship between translators and MT: the attitudes of translators and to translators.

Attitudes to translators

The most obvious change over the last year has been the much greater acceptance in MT circles of the translator's role. A negative translator role - that of obstructing the introduction of MT - has of course long been a possibility, though not always discussed in public. Now, however, it is a positive translator

role which has come to the fore: the contribution of the translator to improvement of an MT system. Translators do know about translation, after all; and whether or not they can express this knowledge in the same ways as academic linguists, there seems now no doubt that translators can contribute to development, whether by supplying feedback from the postediting process, or by specific development work on a system.

Much of the impetus for this change has come from the Commission of the European Communities, to whose generous sponsorship we also owe the ambitious scale of this conference. It was under a Commission study contract that the Cambridge Language Research Unit established that a non-computational translator could decode some SYSTRAN program and suggest improvements. (This was myself, being dragged kicking and screaming into MT - screaming for more, very soon, for it proved a fascinating field.) Over the last year a way has been devised of annotating SYSTRAN so that any post-editor who wishes can read the program and, in all probability, improve it. Higher-quality MT does indeed seem to demand the insight of the professional translator, as well as the skills of linguistics experts and computer scientists. Meanwhile, this year has also seen a pilot experiment on the use of machine translation in the Commission.

This, then, was one recent change. Another was the meeting of SYSTRAN users at La Jolla, California, in April 1981. The first such meeting to include posteditors, its purpose was to co-ordinate the feedback from users, so that linguistic improvements made to one user's SYSTRAN could be transferred to others, whether in the Americas, Europe or Japan. The advantage of linguistic feedback from translators is that translators, besides knowing a lot about how language behaves - or misbehaves - tend to be unusually thorough. For example, it had been predicted that my own work on, as it were, teaching the European Commission's English-French and French-English SYSTRANS to translate patents (1), in making the system better at translating patents, would make it worse at translating other documents. Patents, after all, would appear to combine most of the linguistic abnormalities of both legal and technical texts, and the last thing you want is for your more normal texts to read like patentese. However, all those involved in this development work, inside and outside the Commission, were translators, and of the many hundreds of dictionary or programming changes made under that contract, none, according to the Commission, has yet been found to produce errors elsewhere. Quite the reverse: translations of other texts improved. (A rigorous check was impossible in the absence of quality control procedures such as those of the US Air Force, reported by Dale Bostad at this conference (2). Some degradation must surely have passed unnoticed. Even so, it seems likely that degradation was very far below the 30% found by Wilks.(3))

Attitudes of translators

Parallel to this new interest in translators (not evinced by all linguists and computer scientists, of course, but increasingly widespread), there has been some change of attitude among translators. This is largely, we think, because of the availability of hardware, perhaps particularly word processors. Indeed, this week's "Punch" has a cartoon by Nick of a couple on a sofa at a party: "How fascinating!" says the lady, "Tell me some of the words you process." As more and more translators use wordprocessors or have on-line access to databases such as the term bank Eurodicautom, people are becoming used to the idea of working with machines. And

a different kind of machine aid, the Weidner MT system - which offers word processing and is sensibly marketed as an aid, not a replacement, for translators - has been available for rent here for the last year (ITT will be installing one this month.) All in all, significantly more translators than a year ago have experience of computers, or at least of computer output. And as the new machine aids have begun to be taken for granted, more translators have begun to take a serious interest in machine translation.

Definitions of MT and MAT

What then is this thing, and what can it be expected to do? First of all, how does this conference define machine translation (MT) and machine-aided translation (MAT)? For the purpose of this conference, quite simply, "machine translation" is automatic translation, with or without human assistance. "Machine-aided translation" (at least in the conference papers I have seen) is again automatic translation, but always with human assistance; it is not used, as it sometimes is elsewhere, to mean machine aids for translators (usually excluding automatic translation).

Here is a quick summary of the kinds of machine translation on offer, for the benefit of those who are new to the field.

System types

First of all, we may isolate four approaches, or system types:

- 1) The systems which concentrate on one corpus of text, which they translate more or less well. Another text, even very similar, may throw them. Example: TAUM-Aviation. These so far have been essentially research systems, and not the concern of this conference.
- 2) Systems designed to translate artificially restricted language, either pre-edited natural language or specially written texts in a limited range of syntax and/or vocabulary. Examples are METEO and the Xerox SYSTRAN, about which we shall hear later today, and TITUS at the Institut Textile de France.
- 3) Interactive systems which make it particularly easy for the user to put in his own dictionary and even text-related vocabulary. Examples: Weidner, and CULT in Hong Kong which has translated the Chinese mathematics journals, binding the printout and selling it to foreign libraries. (4)
- 4) The "try anything" system which will have a go at any natural language thrown at it, albeit with varying degrees of success. The obvious example is SYSTRAN.

Modes of MT

Then there are the different modes of MT:

- 1) Pre-edited or specially written input: A pre-editor alters the source text in any of a number of ways before it is machine-translated. He may mark parts of speech, identify the boundaries of clauses or of expressions, try to spot and sort out ambiguities. (I say "try" because we are notoriously unreliable in spotting what is ambiguous to a machine.) To pre-edit effectively is not unlike half translating the document in one's head, and it does not obviate the need for postediting. It may however be

worthwhile if a text is to be translated into more than two target languages. I will also mention much simpler operations, sometimes regarded as pre-editing, performed by the keypunch or other operator who actually inputs the source text if it is not already in machine-readable form: the correction (often unconscious) of obvious errors, the insertion of layout codes and perhaps the marking of words which are not to be translated. As for specially written input, there are various approaches, some of which will be explained in other papers at this conference.

2) Postedited output: The raw machine translation is revised by a posteditor, normally a translator. He may use a word processor or work on paper (known as hard copy). This activity, recognisably comparable to the revision of human translation, and yet very different from this, will be covered in detail in the conference.

3) Interactive editing: The editor intervenes during the machine translation process, perhaps by inputting vocabulary before running the MT program (as suggested earlier with reference to the Weidner), perhaps by helping the machine to analyse correctly or by performing others of the operations mentioned in respect of pre-editors.

4) Raw output: Unedited machine translations, faulty though they are, may be adequate for some purposes.

Translation aspirations

This, to be sure, is the point at which we must ask, what is "adequate", and even, far more fundamentally, what is "a translation"? The professional translator's first reaction to that question must always be that a rose is a rose is a rose - that a translation should be a "true translation", even though he knows that, philosophically speaking, there is no such thing. In fact, a play on in London at present, Brian Friel's "Translations", centres on the impossibility of translation, in that case between Irish and English 150 years ago. Reluctantly, the schoolmaster agrees to teach the heroine English, but he warns her:

"Don't expect too much. I will provide you with the available words and the available grammar. But will that help you to interpret between privacies? I have no idea. But it's all we have."

So we strive for perfection, knowing that it is out of reach.

The philosophy of machine translation is rather different, and I believe it always will be. The machine translation of natural language, unless the text concerned has already been run through the machine and the system changed accordingly, will surely always be imperfect. Natural language is unpredictable, constantly throws up new variations. Not only do people have different styles, different dialects, but language evolves: the man broadcasting the traffic news last week attributed a traffic jam in the Old Kent Road to "damage to a - er - person-hole cover". Then there are the mistakes, of course: printing errors (in my SYSTRAN study a patent for treating depression in a patient claimed "a method of treating depression is a patient" - an error faithfully reproduced by both the keypunch operator and SYSTRAN); punctuation errors (there is a world of difference between the exhortation "Take your wife seriously" and "Take your wife, seriously", as the advertising agents for the Japanese Nikon camera knew when they captioned a redhead);

grammatical errors, particularly in non-native speakers (the hovercraft to Calais lists among its duty-free goods "toilet waters/ eau_de toilettes") ...

So perfection is not something to which the machine can aspire. Is perfection, however, always the most appropriate goal? Where time, translators and money are plentiful, the answer is probably yes. Where any of these is short, the answer may be no, for a full, "true" translation may be neither practicable nor essential. Not all translations are wanted for publication or for legal purposes. Clients often opt for a plain "information" translation done without exhaustive research ("But I only want to know what it says", they say); or they ask for the selective translation of parts of a document, or even simply for the gist of the document. Thus the translation consumer is already offered a wide range of products, and the possibility of machine translation simply adds more.

.... and specifications

Given the variety of kinds of translation required and available, clients will have increasingly to specify the kind of translation needed for a given job. This theme of "translation specifications" was explored in a Translators' Guild seminar in November 1980 (5), and it may be instructive here to examine one proposed series of specifications (Richard Simpkin's): literary, legal, publication, information, selective and gist (or, if more formal, abstract).

MT's place in the translation market

Which of these may be worth machine-translating? Not, of course, literary works, which of all translation types are most concerned with "interpreting between privacies". Literary translation is not the concern of this conference. For legal and publication purposes, on the other hand, MT can sometimes be postedited to a suitably high standard: patent MT, at least of chemical patents, is sometimes adequate for a legal user; and various speakers at this conference machine-translate documents for publication, though not of course documents we regard as "literary". As for documents wanted for information, MT has been used for these for many years, whether in North America, Europe, the Soviet Union or the Far East; even raw machine translation may be useful for information scanning. Raw MT can also assist in "selective" translation, by indicating which passages need a postedited or human translation. "Gist", of course, is a different matter from pure translation, but presumably any MT good enough for information scanning can be expected to assist the person extracting the gist.

Raw MT must of course be used with care. I believe that all raw MT, indeed, all machine-generated text, should be clearly recognisable as such. Ideally, every page should be indelibly printed with the words:

DANGER
MACHINE TRANSLATION
QUICK AND DIRTY!

Seriously, some identification is essential. Again, the term "machine pre-translation" may be more realistic than machine

translation", and less open to abuse.

MT and people

I would like to end with a glance at the likely implications of MT for translators and their clients.

First, two negative possibilities. One, of course, is the replacement of translators by machines. Now, it is noticeable that the translators most worried about MT are the ones with no experience of it. The translators who try to get to grips with it very soon, perhaps in half a day, recognise two things: that MT might work after all; and that it won't work well enough to threaten them. As far as I can see, the only translators who need worry are those who can't do much better than the machine - except, of course, at camouflaging their mistakes! That (often unconscious) camouflaging skill makes them the more dangerous. These are the translators who bring our profession into disrepute, usually undercutting the rest of us, and certainly constituting a danger to the public. I, for one, would not be sorry to see them turn their hands to something else.

For the rest of us, I believe, there will be ample work. Not only have the spread of industrialisation and the "information explosion" produced a huge new need for translation, largely unfilled: the very availability of machine translation for information scanning may generate an interest in large numbers of documents of which present readers are not even aware.

The second negative possibility is the possible malign influence of machine-generated language. Will it affect our style and ultimately - far worse - our perceptions? My position on this has not changed since I was first working on MT in 1979 (6). I still see no real sign of infection, but I believe it a more real risk than unemployment, and if it ever did occur, it would be very grave. I know that many of my colleagues share this concern, and I hope that this conference will shed some light on the matter.

Happily I can end on a more positive note. As Professor Sager has said, machine translation enables us to offer the translation user "a wider range of products". (7) I believe that computers, whether as "translating machines" or machine aids for translators, will both streamline and diversify the translator's job. Both we and our clients will find our views of that job changing. The translator will grow more flexible, more versatile. And the client, seeing the variety of translation specifications to choose from, will become more aware of the translator and his special skills,

REFERENCES

- (1) Lawson, V. Final Report on EEC study contract TH-21 (Feasibility study on the applicability of the Systran system of computer-aided translation to patent texts) (Commission of the European Communities, Luxembourg, 1980).
- (2) Bostad, D. Quality control procedures in modification of the Air Force Russian-English MT system, elsewhere in this volume.

- (3) Wilks, Y. and LATSEC, Inc., Comparative translation quality analysis (Final Report on contract F 33657-77-C-0695).
- (4) Loh, S.-C. and Kong, L. An interactive on-line machine translation system, in: Snell, B.M. (ed.) Proceedings of Aslib seminar Translating and the Computer (North-Holland, Amsterdam, 1979) .
- (5) Drazil, M. and Bacon, J. Guild seminar on translation specifications, Translators' Guild Newsletter 6, 1 (1981) 7-9.
- (6) Lawson, V. Tigers and polar bears, or: Translating and the computer, The Incorporated Linguist 18, 3 (1979) 81-85.
- (7) Sager, J.C. New developments in information technology for interlingual communication, in: Taylor, H. (ed.) Proceedings of Aslib/Translators' Guild conference Machine Aids for Translators (Aslib, London, 1981).