

# A DIFFERENT APPROACH TO MACHINE TRANSLATION

*Karla Déjean Le Feal*

The emergence of data processing has brought in its wake the development of a whole series of more or less fully automated machine translation systems. Although they differ in many respects, they are all based on a linguistic approach to translation.

And yet it is a well-known fact - and indeed one to which human translation bears constant proof - that in most cases this approach is unable to meet the very purpose of translation, which is to communicate a text's cognitive content. The most obvious reason for this is, of course, that languages do not divide up reality in the same way. So the various "meanings" of the words that make them up do not match perfectly from one language to another. Furthermore, in the language (as distinct from its use), such meanings only exist as virtualities. It is only in texts or utterances, in other words, when the language is actually used to communicate, that one of the word's virtual meanings is crystallized and becomes unambiguously clear.

More importantly, the way in which different languages are used to express the same idea is not identical either. Since, contrary to the language itself (which is finite), in the use of language (parole), there is an almost limitless number of potential utterances - that, moreover, do not always comply with the rules on standard usage set up within the language in which they are expressed - there is no way to predict how, in a given context, an idea will be rendered in one language as compared to another.

It is thus apparent that any translation system based on the principle of pairing off words or expressions of different languages can only work at all if it has been specially calibrated for a single referent each time. And even then, only provided this referent is of a type such that the number of possible meanings for each word that are likely to be activated

in the text is kept very small, as well as the range of utterances used to refer to it.

This is the reason why current machine translation systems are only used in certain, very specific fields for which there is sufficient translation turnover to justify the high cost of the computerized systems and to make their shortcomings seem acceptable. For instance, parsing is still a problem, even though the type of text produced in such specific fields tends to be highly standardized. Unless the translations are merely intended for information screening, they still have to be post-edited after being produced by the machine. As this type of post-editing is rather tedious work, and is generally done under pressure of time (seeing that the purpose of fully automated machine translation is to speed up the process), even the finished product tends to "smack" of translation and to lack in clarity. That is why even in such a specific field as the translation of hard-ware and software manuals, IBM France, for instance, does not rely on a fully automated machine translation system, in spite of the high translation turnover and stringent time constraints. As the head of the IBM France translation department, Marcus Dornbusch, put it<sup>1</sup>, none of them was found convincing enough in a field where the consolidation and increase of the manufacturer's market share depend on the quality of the translation of his written material.

Outside the confines of these very narrow, specific subjects, linguistic translation type systems are quite ineffective, no matter how sophisticated they are, if only because the average text has so very many referents and their designation gives rise to such an endless variety of forms of linguistic expression. The only way to translate this type of text is to work out its cognitive content and then re-express this content in keeping with the constraints of the target language.

Alas, these two major operations of translation are unlikely to be feasible by computer for a long time to come, and indeed, perhaps never will be. Which is why, apart from machine translation systems proper, data processing has only been used to provide the translator with faster and better terminology banks.

And yet, data processing could assist translation in another way. It could act as a bridge between the two main phases of translation : i.e. the working out of the cognitive content and its re-expression. This could be done quite simply by integrating current computer technology into the translation process using an approach that would involve the automation only of easily-processed tasks and yet would have a considerable impact on translating both in terms of quantity and quality.

## THE MODEL USED

What a translator means by the working out of a text's cognitive content - including implicit content - implies stripping it of its original verbal packaging. In other words, the translator tries to work back to the author's thought process before it was worded in a given language, because it is this thought that has to be expressed in the target language.

This means that the transition from cognitive analysis to reformulation hinges on a completely non-verbal thought process, which, obviously, cannot be encoded unchanged. So the pattern of operations in normal written translation is not suitable as a model in designing a machine translation system.

Two adjustments have to be made for the computerization of the interlanguage bridge. First of all, instead of being performed just once during the translation process, deverbalization must take place twice : once as the result of the cognitive analysis of the source text and a second time prior to the re-expression of this cognitive content in the target language. But contrary to the pattern of operations in manual translation, deverbalization will not serve the purpose of bridging the language gap, but of reformulating the author's thought process in the source language in such a way that it can be easily code-switched by the computer, and then of rewording this thought in the target language on the basis of the code-switched material.

The second adjustment consists of inserting an additional stage into the translation process, i.e. the recording of the result of the cognitive analysis in the source language. The notes taken in consecutive interpreting provide a suitable model for this recording stage. Indeed, because there is a time lag between the understanding and the re-expression phases of the translation process in consecutive interpretation, the interpreter has to record in written form the outcome of his cognitive analysis. In other words, what he jots down on paper, either in the form of words, or of symbols, is the message<sup>2</sup> worked out from the speech during the listening phase. The layout of the notes on the page replaces syntax ; linguistic elements are used only very sparingly. The level of language of the speech is integrated into the cognitive analysis and is mentally recorded by the interpreter as the stylistic level to be complied within the reformulation.

This approach to note-taking, which is used by professional interpreters everywhere, is systematically taught to students in interpretation, because the layout of the notes on the page and their verbal unclutteredness mean they in no way obscure the cognitive content of the original speech, but rather reflect it particularly clearly. The only major difference between notes taken by different interpreters consists in the specific signs used which are individual and indeed often invented on the spur of the moment, as needed.

In order to make this set of words and signs communicate the initial message to someone who has not read the text ( or heard the speech ), but who has the same ability to understand it, all that is required is to standardize the way the ideas are represented, to indicate the style in which the initial text was written and for what category of readers it was intended, and, without changing the basis approach, to pad out the notes, by adding points that the interpreter does not have time to write down, as his notes are taken at the rate of speech.

To be complete, it should also be mentioned that the interpreter normally takes his notes in the target language, i.e. after interlingual deverbalization. However, his approach to note-taking can just as well be used for recording the cognitive content in the source language, i.e. after intralingual deverbalization. An example is given below :

**Text chosen :** The first paragraph of an article on the situation of people with jobs using languages in the Federal Republic of Germany ( " Language Monthly ", No 25, October 1985 ) ;

**Style :** Standard journalistic style ;

**Category of Readers :** Professional translators and interpreters, language teachers ;

**Explanation of the signs used :**

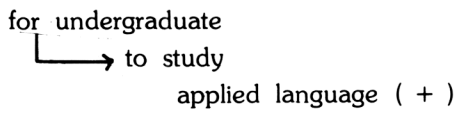
- ∅      currently, at present, now, today, etc.
- ( word )      major, big, great, high, important, etc.  
=====
- ↓      in order to, so as
- ∪      words linked by this sign are to be considered as compound words
- ( + )      plural
- ↳      relative clause
- ∅ effort  
=====
- to be made  
↓ in Federal Republic of Germany
- to unite  
↓ force
- ↓ language teacher ( + )  
↓ and professional linguist ( + )
- to face  
challenge  
language market  
↳ to be characterized  
by unemployment  
=====
- trained } translator  
          } interpreter
- and by low price  
for freelance translation ( + )
- and by gloomy prospect  
for undergraduate  
↳ to study  
applied language ( + )
- } University  
West German }  
  } College

This set of notes, although presented in a very simplified form languagewise, is sufficient to understand the text's cognitive content. The notes are clear due to the way they are laid out on the page, so that even complicated syntactic structures, such as that

sentence, are easy to follow.  
At the same time, this layout allows the notes to be easily code-switched by the computer: There is no need for parsing, verbs do not have to be conjugated, tenses and plurals are only indicated when

the context does not make them obvious and even then only by means of markers that merely have to be copied. While compound words like the "Federal Republic of Germany" are marked by the sign  $\cup$ ,

word clusters like "applied language studies undergraduates", used in the text, are broken down and transformed into, for instance, a relative clause:



The words actually noted are also chosen with a mind to the code-switching operation they will have to undergo without becoming unintelligible. Idiomatic expressions like "efforts are currently under way" figur-

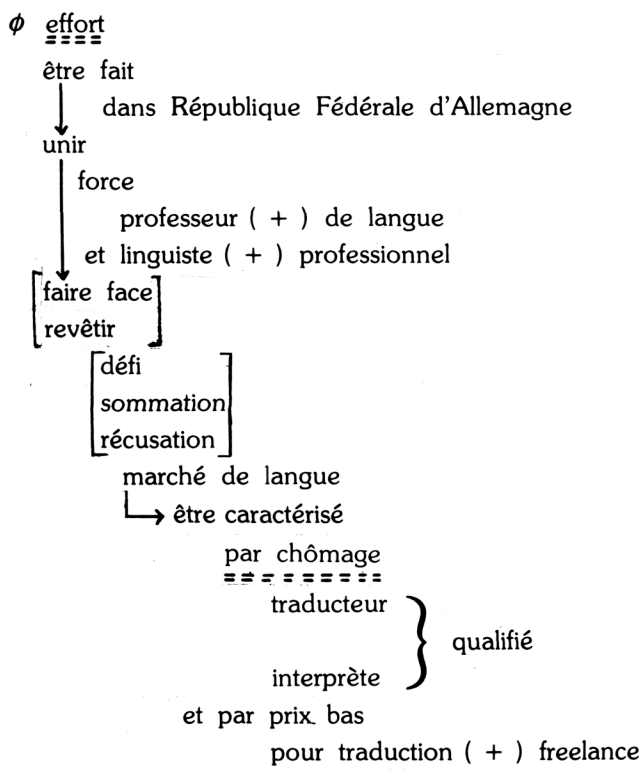
ing in the text should thus be avoided in the notes, which is why this expression has been replaced in our example by.

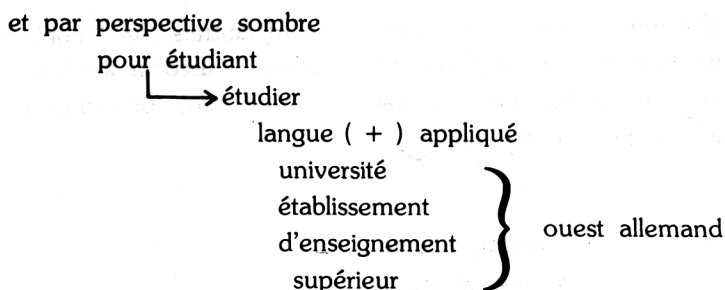
$\phi$  effort  
 to be made

Note also that the word "price" in  
 and by low price  
 for freelance translation ( + )

has replaced the word "rate" which was actually used in the text, but which is highly polysemous. This set of notes can now be easily code-switch-

ed by the computer. The result of this operation<sup>3</sup> is shown below:





As can be seen from this example, code-switching does not make the notes harder to understand. There is some remaining polysemy, but this is not a problem either, because the notes are then given to a human editor for processing, who, being familiar with the subject and aware of the context, will easily pick out the appropriate term in each case. For instance, here it is obvious for a French - speaker

reading the notes that only "faire face" and "défi" are relevant, because they are the only terms that fit into the context and the subject under discussion. And thus, with the note on the style of the text and the intended readers, the human editor is quite capable of producing a text in the target language with the translation shown below obtained using this technique.

Major efforts are currently under way in the Federal Republic of Germany to unite forces of language teachers and professional linguists to face the challenges of a language market characterized by high unemployment among trained translators and interpreters, low rates for freelance translation work and gloomy prospects for applied language studies undergraduates at West German universities and colleges.

Un gros effort est actuellement entrepris en République Fédérale d'Allemagne par les professeurs de langues et les professionnels du langage afin d'unir leurs forces pour faire face à la crise qui touche le marché des langues. En effect, le chômage sévit parmi les traducteurs et interprètes diplômés, les traductions free-lance sont sous-payées, et il ne reste aux étudiants en langues appliquées des établissements de l'enseignement supérieur ouest allemand que de bien sombres perspectives quant à leur avenir professionnel.

The French text is a true translation, because it says exactly the same as the English text in the same level of language, although, or rather because, it is not bound to the form of the original. While French and English are closely related languages, each has its own constraints. Note, for instance, the different ways in which the fact that two words (effort and unemployment) were underlined in the notes has been taken into account by the French post-editor. While in the first instance, we find an adjective ("gros") in the French version, i.e., a language equivalent of "major", in the second case the idea of massive unemployment has been rendered by the verb "sévir", which would not have been used if only a few people had been out of work. This example illustrates that - as was pointed out above - the way in which a given idea in a given context will be expressed in two different languages is not only likely to be different,

but also highly unpredictable. An appropriate reformulation in the target language can be found only if the post - editor deverbilizes the cognitive content of the notes before fitting it into the moulds provided by the target language.

This may also imply that implicit content is made explicit or vice versa, either because of differences in the background knowledge of the intended users of the translation or, simply because of language constraints. It is for this latter reason that the "gloomy prospects for applied language studies undergraduates", for example, have been qualified in the French text by the addition of "quant à leur avenir professionnel", which was only implied in the source text.

Contrary to the post-editing involved in fully automated machine translation systems, which consists of repairing the most glaring mistakes and improper wordings in an otherwise already finished text, the

rewriting of the original text in the target language on the basis of the code-switched notes is not a tedious, but quite a creative and rewarding task. Contamination of the target text by the source language needs hardly to be feared, since it is a purely intralingual exercise and, in most cases, the post-editor will not know the source language.

Moreover, the unfinished form of the notes prompts him to adopt a critical and cautious attitude toward the words suggested to him and to rely primarily on his own judgement. Therefore the system is also quite safe, even in the case of incomplete code-switching. The word "linguist", for instance, had been rendered in the French notes by only one word, i.e. "linguiste" (which means somebody who has studied linguistics), while in English it can also mean

someone who has studied foreign languages. But although this polysemy had not been brought out in the code-switched notes, the post-editor understood whom the term "linguiste ( + )" "professionnel" referred to and used the generic term for translators, interpreters etc., i.e. "les professionnels du langage".

The form of the notes also facilitates the restructuring of the text's cognitive content, whenever this is required either because of differences in the way of presenting ideas in the source and the target culture or because the target text is not meant to have the same function as the source text.

It goes without saying that this machine translation model does not depend on the language processed. An example of a French text translated into English using the same technique is given below :

**Text chosen :** The first paragraph of an editorial in "Le Monde" of 26/27 June 1988 on the illegal storage of toxic industrial waste from Western countries in Africa.

**Style :** carefully written, journalistic style.

**Category of Readers :** educated Frenchmen, intellectuals.

**Explanation of the signs used :**

L→	relative clause
( + )	plural
'φ	just happened
???	direct interrogative clause
( > )	want to, try to, intend to, etc.
↓	in order to, so as
↗↘	increasingly, more and more

**Titre :** AFRIQUE POUBELLE

nation ( + ) industrialisée

L→ 'φ décider

à Toronto

annuler

dette ( + )

pays ( + ) pauvre

??? ( > ) exploiter

↓ pays ( + ) pauvre

économiser

argent

par entreposage  
déchets industriels  
└─→ stockage  
ou traitement  
conformément norme de sécurité  
être ↗ ↗ coûteux

question  
être justifié  
en Afrique  
où  
plusieurs scandale  
ϕ éclater

homme ( + ) d'affaires occidental  
et responsable ( + ) local  
être impliqué  
dans scandale ( + )

As can be seen, the pre-editing has been carried out in the same manner as for the English text, except that compound words ( e.g. norme de sécurité ) are not marked by a sign but by the preposition " de " used only for that purpose.  
It may be worth noting that just like post-editing, pre-editing can also mean making the text's cognitive content more explicit. For instance, in the text the verb corresponding to

??? ( > ) exploiter

is in the future tense. But the future here indicates intention and this is made explicit in the notes by the use of the sign ( > ).  
Also note that pronouns are systematically avoided and replaced by the nouns they stand for.  
The notes are shown below after ( manual ) code-switching into English :

Title : AFRICA DUSTBIN

industrialized nation ( + )  
↳ ϕ to decide  
in Toronto  
to cancel  
debt ( + )  
poor country ( + )  
??? ( > ) [ to exploit  
to run ]  
↓ poor country ( + )  
to save

[money]  
 [silver]  
 by storing  
 industrial waste  
 ↳ storage  
 or [treatment]  
 [salary]  
 in conformity with safety standard  
 ↗↗ to be expensive

Question  
 to be justified  
 in Africa  
 where  
 several scandal  
 'ø to burst

Western businessman ( + )  
 and local person ( + ) in charge

to be [implied]  
 to be [involved]  
 in scandal ( + )

Although there are still a few polysemous words in the notes, they do not reduce the intelligibility of the whole set since the relevant meaning can easily

be deduced from the context by anybody familiar with the subject matter, as can be seen from the English translation below drafted on the basis of these notes :

### L'AFRIQUE POUBELLES

Les nations industrialisées, qui viennent de consentir une remise de dette aux pays pauvres à Toronto, ferontelles par ailleurs des économies sur leur dos en se débarrassant chez eux de déchets industriels dont le stockage ou le traitement, en respectant les normes de sécurité normales, est de plus en plus coûteux? La question est posée en Afrique, où ont éclaté plusieurs scandales qui mettent en cause des hommes d'affaires occidentaux et des responsables locaux.

The English text only rarely uses the same words as those code-switched from the French in the notes. Most of the time, the words in the notes merely act as pointers to the message, which the editor must

### THE AFRICAN DUSTBIN

Are the industrialized countries, after deciding recently in Toronto to write off the debt of the poorest countries, now trying to save money at their expense by dumping their industrial waste on them, as the storage and treatment of this waste in accordance with proper safety standards is becoming increasingly expensive? The question is not unfounded in the case of Africa where a number of scandals, involving Western businessmen and their local contacts, have recently come to light.

then deverbalize before being able to find the appropriate wording in keeping with the constraints of the target language and the style of the text. Note, for instance, that although in the notes, the second sen-



tence of the French text had been broken up into two blocks for clarity's sake, the cognitive content of these two blocks has again been expressed in only one sentence in the English translation.

To sum up, in this approach, the two key operations, i.e. cognitive analysis and re-expression are carried out by human editors, because, for the time being, only human editors are capable of doing them. But, contrary to human translation, these operations are performed within only one language system at a time, since the inter-language bridge is computerized.

## FUTURE PROSPECTS

Although relying only partially on computers, a machine assisted translation system of this type opens up new prospects in inter-cultural communication. First of all, as regards quantity, it helps increase the number of translations possible, especially between and from languages that are little known outside the country or region where they are actually spoken, as the pre- and post-editors do not require any foreign language knowledge at all. Since the inter-language link is computerized, the translation activity per se is replaced by two sets of operations carried out by two separate editors, each working exclusively in his mother tongue.

This means that they need to have only some of the skills normally required to do a translation. In fact, the pre- and post-editing of the text only require thorough knowledge of the subject matter together with analytical ability and, as regards post-editing, with ability to write well. It is obviously much easier to find people with those skills than to find translators, especially - as mentioned above - from little known languages into widely used ones. Indeed, while there may be enough Japanese, for instance, willing to learn English well enough to acquire translation skills, there are not so many native English speakers ready to study Japanese. Yet, translations into English cannot be done by Japanese. They have to be done by native English speakers. But while only few native English speakers may be inclined to learn that language, there should be no dearth of post-editors, especially as, with remote transmission, the pre- and post-editing can be carried out in the countries of the source and target language respectively.

This splitting up of the translation process and

the concomitant division of skills and better use of local human resources should also provide translations that are just as good, if not better, than many manual translations currently available. The main reason why many of these are unsatisfactory is that the translator either does not know the source language or the subject matter well enough, or has no writing skills or has not been trained in the translation technique. This in turn is often due to the over-simplistic idea most people have of translating, while it is in fact a very demanding task. It seems quite likely that if the operations that still have to be carried out by human editors are less complex and better understood, in particular because they can no longer be considered as the mere application of linguistic knowledge, the general view held of these activities and the qualifications they really require may be sufficiently well matched up to provide satisfactory standards.

The second reason why translations produced by this system are likely to be of better quality than manual translations is that the replacement of the translation process by two sets of intralingual operations eliminates the most difficult stage of the translation process, i.e. complete deverbilization. There is, indeed, a difference between the degree of deverbilization necessary for intralingual and interlingual processing. While the former requires deverbilization only up to a certain extent, to which we normally and unconsciously deverbilize when reading, for instance, the latter has to go through a completely non-verbal stage which can only be attained by applying a specific technique. Circumvention of this hazardous stage therefore leads to a simplification of the translation procedure and should improve its results.

It goes without saying - as indeed the choice of texts shown above indicates - that this machine translation system can be used for all types of texts that fall within the scope of functional translation. The simplification of the code-switching operation, thanks, in particular, to the representation of syntax by the note layout and the related reduction of the basic vocabulary make it a very inexpensive system with regard to development costs, as well as to the hardware necessary to run it. The storage capacity it requires does not exceed that available in most PCs.

*In conclusion, the system put forward is akin to human translation in technique and to machine*

translation by its use of data-processing as a bridge between two languages. As such, it has a role to play in supplementing both. It will probably never break records in translating speed nor beat the lowest cost per page translated, but should nonetheless help

promote communication across language and cultural barriers by providing high-standard translations that will not undermine the distinct identity and integrity of the target language.

### FOOTNOTES

- <sup>1</sup> "Les Cornes de Moïse", in *Le Français dans le Monde*, Numéro Spécial, Août - Septembre 1987.
- <sup>2</sup> Used here in the sense of cognitive content.
- <sup>3</sup> Pending the development of the necessary soft-ware, the code-switching operation was carried out manually.