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, to communicate a text cognitive content. The most obvious reason for this is, of course, that languages do not divide up reality in the same way, and so the various "meanings" of words do not match perfectly from one language to another. Furthermore, in a language (as distinct from its use), such meanings only exist as virtualities. It is only in texts or utterances, when the language is actually used to communicate, that one of the word virtual meanings is crystallized and thus 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 a language itself (which is finite), the use of a language (parole), entails an almost limitless number of potential utterances - that, moreover, do not always comply with the rules on standard usage - 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 if it has been specially calibrated for a single subject each time. And even then, only provided this subject is of a type such that in related texts, the number of possible meanings for each word that may be activated, as well as the range of utterances likely to occur, remain quite limited.

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 computerized systems and to render their shortcomings tolerable. For instance, parsing still poses 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 time pressure, precisely because the purpose of fully automated machine translation is to minimize the translation time, even the finished product tends to "smack" of translation. That is why even in such a specific field as the translation of hard-ware and soft-ware manuals, IBM France, for example, does not rely on a machine translation system in spite of the high translation turnover and of stringent time constraints. Indeed, as Marcus Dornbusch, the head of the IBM France translation department, put it, none of them was found convincing enough in a field where consolidation and increase of manufacturers' market share depend on the quality of the translation of their written material¹.

Outside the confines of this very narrow subject, translation systems of the linguistic type are quite ineffective, no matter how sophisticated

---

they are, if only because the average text has so very many referents whose 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, keeping within 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. This is why, apart from fully automated translation systems, 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: the elaboration 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 which would involve the automation of only easily-processed tasks, but would still have a considerable impact both on the quality and the quantity of translations.

The Model

What a translator means by the elaboration of a text's cognitive content — including implicit content — is stripping the text of its original verbal packaging. The translator tries to work back to the author's thought process before it was worded in a given language, because it is his thought at this non-verbalized stage that has to be expressed in the target language.

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

Two adjustments have to be made for the computerization of the inter-language 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 the second time prior to the re-expression of this cognitive content in the target language. Contrary to the pattern of operations in manual translations, deverbalization will not serve the purpose of bridging the language gap. Rather, the first deverbalization is necessary in order to re-formulate the author's thought process in the source language in such a way that it can be easily code-switched by the computer, while the second is required before re-wording 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: the recording of the result of the cognitive analysis in the source language. The notes taken in consecutive interpretation provide a suitable mode 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. What he jots down on paper, either in the form of words, or of symbols, is the message\(^2\) 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 attain in the re-formulation.

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

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

\(^2\) Used here in the sense of cognitive content.
An interpreter normally takes his notes in the target languages, after an inter-lingual deverbialization. This same approach to note-taking can just as well be used to record the cognitive content in the source language after intra-lingual deverbialization. An example is given below:

**Text chosen:** The first paragraph of an article on the foreign language job market 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 symbols used:**

- 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
- translator
- trained
- interpreter
- and by low price for freelance translation (+)
- and by gloomy prospect for undergraduate
- to study
- applied language (+)
- University
- West Germany
- College

- currently, at present, now, today, etc.
- major, big, great, high, important, etc.
- in order to, so as to
- words linked by this symbol are considered compound words
- (+) plural
- relative clause
- eliminates repetition of two elements linked to the same term
- [] indicates multiple meanings
This set of notes, although presented in a very simplified linguistic form is sufficient to understand the text cognitive content. The notes are clear thanks to their layout on the page so that even sentences with as complicated a syntactic structure as this one are easy to follow.

At the same time, this layout allows tenses 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 to 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 
, word clusters like "applied language studies undergraduates", used in the text, are broken down and transformed into, for instance, a relative clause:

\[
\text{for undergraduate} \quad \downarrow \quad \text{to study} \\
\text{applied language(+)}
\]

The words actually noted are also chosen with the code-switching operation they will have to undergo in mind. Idiomatic expressions in the text, such as "efforts are currently under way" are thus avoided in the notes and replaced by:

\[
\downarrow \text{effort} \\
\text{to be made}
\]

Note also that the word "price" in

\[
\text{and by low price for freelance(+) translation(+)}
\]

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-switched by the computer. The result of this operation is shown below (pending the development of the necessary soft-ware, the code-switching operation was carried out manually):

\[
\downarrow \text{effort} \\
\text{être fait} \\
\text{dans République Fédérale d'Allemagne} \\
\text{unir} \\
\text{force} \\
\text{professeur (+) de langue} \\
\text{et linguiste (+) professionnel} \\
\text{faire face} \\
\text{revêtir} \\
\text{[défi} \\
\text{sommation} \\
\text[récusation] \\
\text{marché de langue} \\
\text{être caractérisé} \\
\text{par chômage} \\
\{\text{traducteur} \}
\text{qualifié} \\
\text{interprète} \\
\text{et par prix bas}
\]
As can be seen from the 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. Being familiar with the subject and aware of the context, this editor will readily select 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 the context and the subject under discussion. And thus, with the note on the style of the text and the intended audience, the human editor is quite capable of producing a text in the target language with the same cognitive content and style as the original, as can be seen from 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 interpreter, low rates for freelance translation work and gloomy prospects from 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 les marchés des langues. En effet, le chômage sévit parmi les traducteurs et les interprètes diplômés, les traductions freelance 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 thing 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 and adjective "gros" in the French version (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 deverbalizes the cognitive content of the notes before fitting it into the moulds provided by the target language.

This process may also imply that implicit
content is made explicit or vice versa, either because of differences in the background knowledge of the intended audience or simply because of language constraints. It is for this latter reason that "glumly prospects for applied language studies undergraduates", for example, has 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 correcting glaring mistakes and improper wordings in an otherwise 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 is minimal since the exercise is purely intralingual and, in most cases, the post-editor will know the source language.

Moreover, the unfinished form of the notes prompts him to adopt a critical and cautious attitude toward the words generated by the computer 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 example, had been rendered in the French notes by only one word, "linguiste" (which means someone who has studied linguistics), while in English it can also mean someone who has studied foreign languages. Even though 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., "les professionnels du langage".

The form of the notes also facilitates the restructuring of the text's cognitive content. This may be required either because of differences in the presentation of ideas in the source and the target cultures or because the target text is not meant to have the same function as the source text.

Obviously, 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 symbols used:**

- relative clause
- plural
- just happened
- direct interrogative clause
- want to, try to, intend to, etc.
- in order to, so as to
- increasingly, more and more
- indicates multiple meanings.

---

**Titre:**

**AFFRIQUE POUBELLE**

- nation (+) industrialisé
- décider
- à Toronto
- annuler
- dette (+)
- pays (+) pauvre
- (+) exploiter
As can be seen, the pre-editing has been carried out in the same manner as for the first example, except that compound words (e.g. norme de sécurité) are not marked by a symbol but by the preposition 'de' used only for that purpose.

Pre-editing, like post-editing can mean making the text cognitive content more explicit. For instance, in the text, the verb corresponding to ???? (_) expléter is in the future tense. However, the future here indicates intention and this is made explicit in the notes by the use of the sign (>). Also, 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 GARBAGE CAN

Industrialized nation (+)
| 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

[ ] 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 since the relevant meaning can be readily deduced from the context by anyone familiar with the subject matter, as can be seen from this English translation drafted on the basis of these notes:
L'Afrique Poubelle

Les nations industrialisées, qui viennent de consentir une remise de dette aux pays pauvres à Toronto, feront-elles 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.

"Trashing" Africa

Are the industrialized nations, after deciding recently in Toronto to write off the debt of the poorest countries, now trying to save money at their expense by dumping industrial waste on them since 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 local officials, have recently come to light.

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 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. For instance, although in the notes, the second sentence of the source 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.

In summary, the two key operations of this approach, cognitive analysis and re-expression, are carried out by human editors, because, for the time being, only human editors are capable of doing them. Contrary to human translation, however, these operations are performed one language at a time, since the inter-language bridge is computerized.

Prospects for the Future

Although it only partially relies on computers, a machine-assisted translation system of this type opens up new prospects in inter-cultural communication. First of all, such a system would 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, given that the pre- and post-editing does not require any knowledge of a foreign language. 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 only need to have some of the skills normally required to do a translation. In fact, the pre- and post-editing of the text only requires thorough knowledge of the subject matter, analytical ability and, in the case of post-editing, the ability to write well. It is obviously much easier to find people with these skills than to find translators, especially from little-known languages into widely-used ones. Indeed, while there may be enough Japanese, for example, willing to learn English well enough to acquire translation skills, there are not as 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. Although few native English speakers may be capable of translating from Japanese, there should be no dearth of post-editors, especially since, with remote transmission, the pre- and post-editing can be carried out in the countries of the source and target language respectively.

Splitting up the translation process with the accompanying division of skills and better use of local human resources should also provide translations that are just as good, if not better, than many anal translations currently available. Current translations are often unsatisfactory because the translator does not master the source language, is not familiar with the subject matter, lacks writing skills, or has not been properly trained in the technique of translation. These shortcomings in turn are often due to the overly simplistic view most people have of translation, when 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 actually require may be sufficiently well matched to provide satisfactory results.

Another reason why translations produced by this system are likely to be of better quality than manual translators is that the replacement of the translation process by two sets of intralingual operations eliminates the most difficult stage: complete deverbalization. Indeed, there is a difference between the degree of deverbalization necessary for intra-lingual and inter-lingual processing. While the former requires deverbalization only up to a certain extent, (the extent to which we normally and unconsciously deverbalize when reading, for example), 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.

The machine translation system presented here can be used for all types of texts that fall within the scope of functional translation. The simplification of the code-switching operation, thanks to the representation of syntax by the note layout and the related reduction of the vocabulary make it a very inexpensive system in terms of development costs as well as hardware. The storage capacity it requires does not exceed that available in most PCs.

**Conclusion**

The proposed system is akin to human translation by its technique and to machine translation by its use of data-processing as a bridge between two languages. As such, it as 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 quality translations that will not undermine the distinct identity and integrity of the target language.