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 [1], 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 test 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 [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 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 :

ϕ 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

ϕ 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 (+)
West German { University / 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 ‿ , word clusters like " applied language studies undergraduates ", used in the text, are broken down and transformed into, for instance, a relative clause :

for undergraduate
└→ to study
applied language (+)

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

ϕ 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-switched by the computer. The result of this operation[3] is shown below :

ϕ effort
être fait
↓ dans République Fédérale d'Allemagne
unir
↓ force
professeur (+) de langue
et linguiste (+) professionnel
[faire face / revêtir]
[défi / sommation / récusation]
marché de langue
└→ être caractérisé
par chômage
traducteur } qualifié
interprète }
et par prix bas
pour traduction (+) freelance