

# Some reflections on the teaching of CAT

Pascaline Merten

Institut supérieur de traducteurs et interprètes,

Haute Ecole de Bruxelles

[pmerten@compuserve.com](mailto:pmerten@compuserve.com)

## Synopsis

Information processing is both a tool for the professional translator and an area of interest to translators. This implies 2 types of teaching for data processing: knowledge of the field and know-how.

The use of machine aids to translation has to respect 2 fundamental principles : the techniques are to be used in the service of content and the document has to be thought as electronic.

As more and more translators are called upon to use MT systems or MAT tools, as those techniques lead to new translation professions, we have to train the students. But an intelligent use of these tools requires some expertise in computational linguistics, which, incidentally, provides an initiation to an intellectual way of thinking that can only be profitable to students in their professional careers.

## I The problems involved

During a recent examination, I asked students to type their answers using their word processors and to structure them by means of "heading" styles. On examining one student's diskette, I found a text correctly presented but apparently without styles. Taking a closer look, I realised that she had, in fact, applied styles, but to her examples, which had become "Heading 1", "Heading 2". Her titles, on the other hand, had been left in the "Body of the text". It took her a while to understand why I failed her for this piece of work, no doubt convinced only of the severity of her maniac teacher. Since then, I start every academic year by explaining to my students two fundamental principles:

- I expect them to use the techniques they learn in the service of content. It is not enough to prove that they know which buttons to push. Information processing in the curriculum of translators is not an end in itself but a means.

- It is absolutely crucial that they think of the document as electronic: the aim is no longer simply to see the document printed, their computer is not merely a typewriter. Of course, if the paper version looks attractive, so much the better. But the final document can now be seen as part of a production process and, even in a task as simple as word processing, it must be remembered that the electronic version will be used more often than the hard copy.

The question fundamental to all teaching is that of the objectives being pursued, since these determine contents and methods alike.

My personal aim is, together with my colleagues, to train translators capable of functioning immediately in any field and of using all the techniques at their disposal both now and in the future.

This implies two types of teaching for data processing:

- of knowledge: knowledge of the field,
- of know-how: the use of appropriate tools.

## II. Theoretical knowledge of the field

There is an ever-growing literature on the subject of data processing, which is a major area of interest to translators. And it is part of our tradition to provide students with basic training in the general fields in which they will work: economics, law, the sciences, etc. Data processing is part and parcel of these fields, in which even a little knowledge will enable them to apply the first principle of Etienne Dolet: *"Il faut que le traducteur entende parfaitement le sens et la matière de l'auteur qu'il traduit"*<sup>1</sup>.

Of course, we are not trying to train programmers, any more than we are trying to train lawyers or economists. Our aim is to convey basic concepts and "to teach students how to learn". They need to understand the text sufficiently to be able to identify what they do not understand. When I hear someone speaking Japanese, I have no idea whatsoever of what is being said. When I hear German, Spanish or Italian, I understand some sentences, while others remain quite unintelligible, because I recognise neither the syntax nor the vocabulary.

---

<sup>1</sup> "The translator must arrive at a perfect understanding of the meaning and subject matter of the author (s)he is translating."

When I hear Dutch, I have no problem identifying the words that I do not understand, because I have a good knowledge of this language and its structures. Ultimately, I can understand the sentence as a whole, even though I do not understand all the words, because I can deduce the sense from the context. When you read philosophy, you must find yourself thinking "I can analyse this sentence grammatically, I recognise every word in it, and yet I don't understand a thing!" Your problem lies at the conceptual not the linguistic level. What is important, therefore, is that the student should understand the subject of the text, the topic of each sentence, because if he manages to identify the problem areas, he will be able to document himself and fill in the gaps in his knowledge.

Of course, you may argue that we are overwhelmed by data processing, everybody has some background in the field. As a practising teacher, I can confirm that the young people arriving in higher education are all familiar with computers to some extent. Even so, most of them have only shreds of information. They have no structured, organised knowledge.

## III Technical know-how

### A. The loneliness of the translator

Besides knowledge there is know-how. Students must be able to use a certain number of tools. Here too, the question of objectives arises. The translator must be able to work in a whole range of different environments, and to do so at the drop of a hat. As a free-lance, in a translation bureau, in a translation department, and so on. In any case, what we have to foresee is the worst-case scenario: having to face the problems posed by that marvellous little machine all by yourself. Even when there are programmers in house, or more experienced colleagues around, there are usually far too few of them to cope with the demand or they are simply unavailable at the precise moment they are needed. This means, if it has not been done already, teaching the operating system, its management and its configuration. I have no wish to see those people who always save their documents in the same place (in the "My Documents" directory or on their diskette) and who are unable, when the need arises, to recover the document they have saved from a CD or the Internet. Those working in Russian must be able to configure their keyboards appropriately, and without help, since the standard installation of Windows does not allow for multilingualism.

Finally, developments in CAT are moving in the direction of network access to the various tools necessary. This, in turn, implies an initiation to networks.

## B. Machine aids to translation and machine-assisted translation

Where software is concerned, I usually draw a distinction between

- machine aids to translation (MA),
- machine-aided translation and machine translation (MAT).

In the first case, the computer provides support for the translator; in the second, it intervenes in the translation process. This it does by automating either the translator's "memory" function by way of a translation memory, or the "translation" function itself by means of a machine translation system. We may include here products that have not yet been marketed or are still in the development stage, like translation checkers<sup>2</sup> or style checkers.

## C. Translation vs. translated document

It is far from clear what should be included in the first category of machine aids. I wish to limit myself to machine aids in the strict sense: word processors, terminological retrieval and management tools, document retrieval engines (either on the Internet or on CD). But things are evolving in such a way that the translator (in fact, anyone producing a document) has to be able to intervene at every step of the production process and, in particular, provide not only the translation on paper but the translated document.

Intervening at every step presupposes that the translator is capable of receiving and sending a file electronically, capable of handling OCR.

The "translated document" can be almost anything: not simply a sheet of paper but a Web page, a document with page layout, or a spreadsheet. Does this mean to say that we have to train students to create Web pages, to use Exell and PageMaker?

## D. Of buttons and concepts

There can be no doubt that in ten years' time the level of the students arriving in higher education will be good enough for us to be able to broaden their training, but I think

---

<sup>2</sup>Macklovitch 1995

that, at the moment, spreading the load too widely could be dangerous. Each tool taken in isolation is easy enough to use; the difficulty lies in orchestrating the use of the whole set of tools, in which there is a risk of losing sight of the essentials. It is better to take a few important tools and highlight what they can do and how they work.

Understanding basic principles is far more important than being able to operate a multitude of tools. I do not believe in "push-button" knowledge. You know as well as I do that a good cook is not someone who is obliged to follow a step-by-step recipe to produce a tasty dish: no, it is someone who draws inspiration from a recipe and then adds a personal touch because they understand the rules of cookery.

As interfaces become increasingly standardised, understanding principles allows users to familiarise themselves quickly with other tools. What matters is understanding the functions of the right-hand button on the mouse, realising the difference between "save" and "save as", recognising, under a variety of appellations, the function that allows you to retrieve an article or a terminological entry, understanding the principles of document retrieval.

Here too, everything depends on the purposes of the curriculum. Do you want to train a passive user of tools or someone who can function actively in the world of MAT? In certain cases, simple information is enough, in which case three things need stressing:

- what the purpose of the tool is and what you can expect from it;
- how you call on-line help and how you use it;
- how you cancel the action you have just made.

In other cases, e.g. word processing, in-depth knowledge is essential in order to be able to use the tool intelligently. Whatever your choice, try not to underemploy tools. People are constantly being caught out by certain procedures because they have no clear understanding of what is going on and they cannot find solutions to the simplest problems. When someone sitting in front of me changes the size of the characters to make the text look bigger on the screen, I see red.

## **E. Knowledge of the field and the principles**

User-friendly tools give the illusion that you can turn yourself into a designer, a musician or an architect at will; nonetheless, there comes a stage at which it is hard to get by without some knowledge of the field you are working in. You do not learn layout or

bookkeeping simply by using PageMaker or Exell. Exell is perfect for basic arithmetic functions, but if you are into statistics, it is worth knowing, for example, what a standard deviation is.

Terminology tools are exemplary in this respect. Consulting a terminological DB does not generally require any substantial previous knowledge, except for some expertise in document retrieval techniques if you want your research to be successful. Creating terminology entries, however, means understanding the principles of terminology: e.g. avoiding creating distinct entries for a given term in different languages or for synonyms. A third level of use is required in designing the data base itself. Even though it is not imperative to know data bases in great depth, some basic understanding of DB is nevertheless useful. A knowledge of terminology, on the other hand, is absolutely essential.

## IV. Machine-aided translation and computational linguistics

### A. Machine translation and post-editing

The translation of the Kenneth Starr report performed by Systran is certainly pleasing to read. However, the interesting thing about using this kind of tool is to come to recognise its limitations, to understand why machine translation often fails to handle certain difficulties correctly, leaving problems which the translator / reviser will have to solve.

Word > le mot
---------------

Here we have an example of ambiguity between common and proper noun. "Word" must be coded as a proper noun with a capital, separately from the common noun "word".

The Director of Oval Office Operations > Directeur des opérations ovales du bureau
---

Two analyses are possible, represented here by dependency trees:

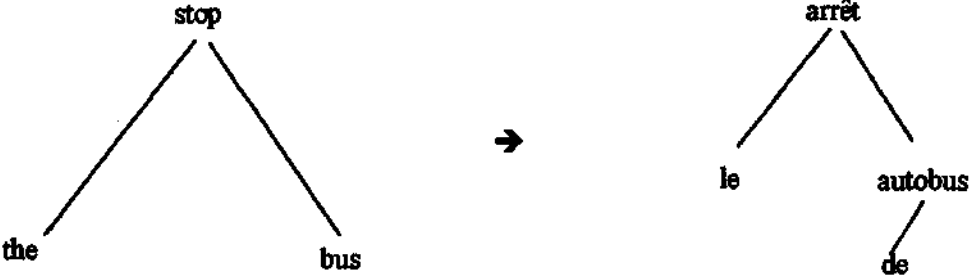
3 spoons butter >  
3 beurre de cuiller

This kind of error is extremely common in machine translations of recipes. To understand the reason, and possibly to find a solution, you need to examine how the system works.

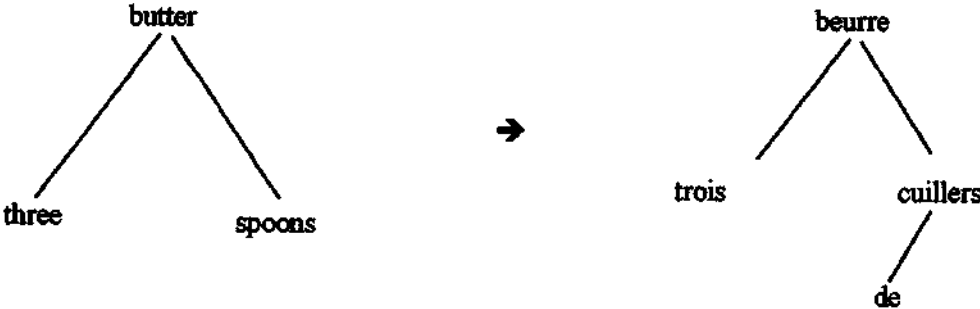
The general translation rule can be schematised as follows:

*The bus stop -> l'arrêt d'autobus.*

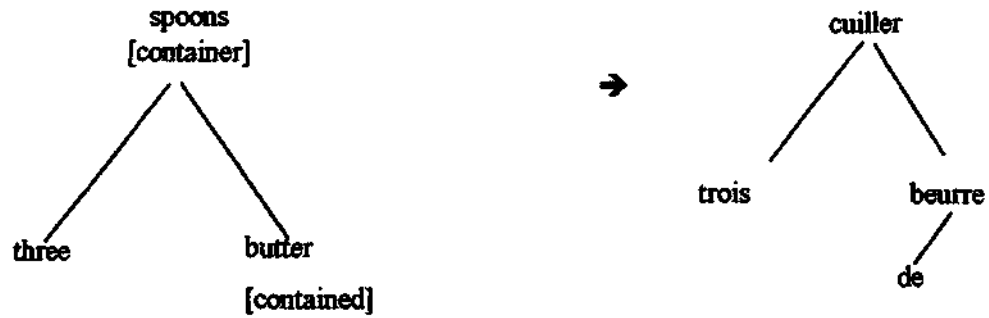
In English, the characterising word (the modifier) precedes the one characterised (the head); in French, the head element comes first.



Applying this rule, the translation of "3 spoons butter" by "3 beurre de cuillers" is logical.



In fact, there is a particular rule that constitutes an exception to the general rule:



Depending on the system you use, you will have to leave the reviser to resolve this kind of problem or integrate the particular rule into the set of rules the system uses.

Conceptualising the problem requires some knowledge of formal linguistics, and finding a solution is made easier by having a basic understanding of computational linguistics. To comprehend, insert and create transfer rules is not a straightforward task, either for a translator, or for a programmer who may never have had any dealings with natural language processing.

In conclusion, whether the system allows the user to enrich only the dictionary or the rules as well, at a certain level a certain linguistic reflection is necessary, one that goes beyond mere technical use of the tool itself.

For revising purposes, a whole set of procedures will need to be brought into play: style sheets for formatting and layout, macros, advanced search and replace functions to facilitate the task.

The examples that follow are drawn from working with MS Word. It would be far easier, of course, to use a language such as AWK, Omnimark or Perl - that is to say a language endowed with real pattern-matching functions - to perform this kind of operation, but we wanted to present realistic examples of translation system practice.

1 spoon butter > 1 beurre de cuiller > 1 cuiller de beurre
--

It will be necessary to write a macro that transforms the sequence "1 2 3 4 5" into "4 5 3 2".



Example > Example (bold) + style "example"

Highlighting examples is done by replacing the expression "Example:" by the same expression in bold and by applying an "Example" style.

Instruction-macro > macro

Instruction-macros > macros

A search and replace of "instruction-macro" by "macro" is made. The plural presents no problem. However, the determiner also has to be taken into account: "l'instruction-macro" becomes "la macro", whereas in the plural, the determiner does not change: "les/des instruction-macros" > "les/des macros".

transposer > inverser

permuter > échanger

Spontaneously, the replacement would be carried out on the basis of the radicals: transpos > invers. In the second example, care must be taken to respect the morphology of French in "échange", "échangeons", "échangeaient". Accordingly, it is necessary to proceed in two steps:

1. permute > échange
2. permut > échange.

Solution > remède

"Solution" has to be replaced by "remède" in the singular and plural forms. To avoid replacing "solution" as part of a word, as in "dissolution", an indication will be given that "solution" is the beginning of the word: "<solution" > remède".

Labor/labour > work

To take such variability into account, it is necessary to use the "generic" characters of Word: "labo\*r > work", which will retrieve both "labor" and "labour".

## B. Translation memories

A translation memory such as the Translator's Workbench of Trados is very easy to use once you understand what a document template is, how to install a template and modify the tool bars in your word processor.

## C. Alignment tools

Alignment tools are still based on text layout. Recognition of correspondences relies largely on format and paragraphs and sentences boundaries. Unfortunately. While waiting for commercial products based on more sophisticated algorithms to make their appearance, we shall just have to accept the fact and process texts to be aligned in advance, in such a way as to guarantee the reliability of the alignment. Here too, we are stepping beyond any simplistic use of the word processor to a level calling for the utmost rigour in using styles.

Once again, it is clear that the document must be thought of as one element in a production process, in which it is not the paper but the electronic product that is being handled.

## D. The new translation professions

There is one fundamental criticism of all this: is it really necessary to train future translators to use a machine translation or machine-aided translation system? I think it is. First, because an increasing number of translators working in big translation departments or bureaux are called upon to use them. Second, because the use of such tools leads to new posts being created: the reviser or translation assistant who prepares translation memories, who loads the terminological DBs and other dictionaries. They sometimes even have a role to play in developing these tools.

As users, they must be able to understand the problems that arise and be able to solve them. However, as has been shown, this requires more than just a little technical know-how.

## V. Methods

Once the decision about what to teach has been taken, there remains the question of how to teach it: of methods.

There is a fundamental difference of attitude between the linguist and the translator. The first commandment of Louis Truffaut is "*Linguistique et traduction tu distingues*"<sup>3</sup>. The linguist observes the language to discover the generalities necessary to the task of formulating theories. The translator has an empirical perspective, deals with concrete cases, with the uniqueness of use.

It is also true, that a competent and experienced translator "doesn't have to puzzle over grammatical classes of words, word order, or length of sentences"<sup>4</sup>. The translator translates automatically, without thinking about the linguistic structures. It's the purpose of computational linguistics to highlight the rules, the process behind this automatic human translation.

Now that computers have become part of the translator's stock-in-trade, the way of working, even the way of thinking about it, is changing. It should now be obvious that incorporating a data processing course into the curriculum of translators is no easy matter. I certainly have still not found the solution. However, two paths seem to me to be unavoidable:

- Combine training in translation with training in data processing. Do not leave your students in the hands of a pure, dyed-in-the-wool "programmer", who will be far removed from the needs of his public.
- Proceed by induction. Students are generally unattracted to theories that appear too abstract to be useful. Start from concrete cases and, if possible, funny ones, because play is an excellent stimulus to learning at any age.

---

<sup>3</sup> "Thou shalt distinguish between linguistics and translation."

<sup>4</sup> Nida 1996: 109

## VI Findings

What the inexperienced student finds difficult is not using each individual tool but orchestrating the whole set of these tools. This poses, in fact, the basic question in data processing: I have a problem; how will I solve it, and with which tools?

Good training in data processing for translators is not limited to teaching them how to use the tools on the market; they must be able to make full use of these to become more effective in the work they do as translators, writers or terminologists.

An intelligent use of these tools requires some expertise in computational linguistics, which, incidentally, provides an initiation to an intellectual way of thinking that can only be profitable to students in their professional careers.

I also consider MA and MAT to be part of the translator's general culture. From this point of view, part of my course can be seen as an extension to the "Background to translation studies" course or, if you prefer, as the "present-day history" of translation.

## VII. Bibliography

- Blampain D., Petrusa Ph. and Van Campenhoudt M. 1991. "A la recherche d'écosystèmes terminologiques", in *Actes du colloque "L'environnement traductionnel"*, Mons.
- Brace C., Vasconcellos M. and Chris L. 1995. "MT Users and Usage: Europe and the Americas" in *Proceedings of MT Summit V*, Luxembourg.
- Clas A. and Bouillon P., eds. 1993. *TA-TAO: recherches de pointe et applications immédiates. Actes du colloque de Montréal*. Beyrouth/Montréal, FMA/Aupelf-Uref
- Guislain G. 1990. *Didactique et communication*, Labor, Brussels
- Hutchins J. 1996. "Computer-based translation systems and tools", in *ELRA Newsletter*, vol. 1 no. 4.
- Hutchins J. 1986. *Machine Translation : Past, Present, Future*, Ellis Horwood Limited, Chichester.
- Hutchins W. J. and Somers H. L. 1992. *An Introduction to Machine Translation*, Academic Press, London.
- Macklovitch E. 1995. "TransCheck - or the Automatic Validation of Human Translations", in *Proceedings of MT Summit V*, Luxembourg.
- Merten P. 1992. "Apport des relations notionnelles à la description terminologique" in *TAMA 92*, Avignon.
- Nida, E. A. 1996. *The Sociolinguistics of Interlingual Communication*, Brussels, Les Editions du Hazard.
- Schubert K. 1997. "Remote Access Translation Services: Software Design with the User in Focus", in *Translating and the Computer 19*, Aslib, London
- Trufiaut L. 1997. *Traducteur tu seras*, Brussels, Les Editions du Hazard.
- Vasconcellos M. 1987. "Post-editing on-screen: machine translation from Spanish into English" in *Translating and the Computer 8*, Aslib, London