

**Harald H. Zimmermann, SOFTEX, Saarbrücken**

## **Machine translation as a practical tool in information processing.**

### **1. Introduction**

In school, I was very impressed by the story of Daedalus and Ikarus. Ikarus lost his life when trying to fly closer to the sun, even though his father warned him of flying too high.

Everybody knows the pictures of men trying in vain to fly like a bird, and even today where hang gliding is a common form of sports only a small group of people is really practicing it.

On the other hand, every year, millions of people nowadays use airplanes to reach places all over the world in a very short time, thanks to the engineering mind of thousands of people working in the big business of aircraft industry.

What are the main reasons to reach such an aim? How can we proceed to succeed? These questions will be made more precise and will be answered by some theses.

### **2. Basic questions and answers**

Everybody knows that translation of natural language is an intellectual process which relies not only on formal identification of words and of sentence structures but on understanding the meaning of a word (word semantics) and of the contextual situation, as well as of the so-called world knowledge.

It is quite clear that there is no MT system available which is able to fulfill this aim of complete "understanding" in its totality, but it is also evident that no single human being is able to satisfy this aim.

Therefore, two basic direction of development and processing in machine translation (further: MT) can be seen:

- (1) replacement of human language processing by machine processes;
- (2) cooperation between human and machine functions in translation.

The first solution is needed in all cases where human activity is not available. In general, this seems to be the biggest market (with the exception, that one day mankind will use only one natural language or at least be able to communicate in one language, e.g. English).

## Cooperation between man and machine

The second solution seems to be the best strategy to reach the highest possible quality, especially if the "forces" of both partners are used and the weaknesses are avoided. At the moment (in 1995), there is no doubt that the usage of machines in translation industry is increasing. This includes the application of electronic glossaries and terminology, word processing systems during the creation of human translations and during post-editing. It will include translation memory techniques as well if this reduces the amount of work.

Therefore, if we speak about "cooperation" between man and machine in MT, we have to think about "higher" degrees of machine application.

This next step is the use of "full" machine translation capacity to provide a so-called "raw" output which then is used for human post-editing. There, also, is a simple calculation (which is well-known in translation industry): If this function will reduce the (total) costs of translation and/or will spare processing time (see the LOGOS' slogan "time to market"), there will be a market for providers of such systems.

Nowadays - and this is very different compared to earlier times - computer processing time, computer space and even the costs of purchasing and licensing an MT system available on the market plays no important role in this market segment. There are several basic questions which have to be answered by the system providers:

- (1) What is the amount of work the customer has to invest to adapt the system to his application environment? - This component, for instance, has to be compared with the amount of work for instructing human translators as newcomers.
- (2) What is the amount of continuous work customers will have to earmark to this part of the translation processing? - This, for instance, could be compared with the amount needed for human terminology work.
- (3) Is there a possibility (of the system) to overcome - step by step - stupid MT solutions or has the human user (as post-editor) to be a "slave" of the system all the time? - If there is no sufficient answer, the consumer acceptance will be very low.
- (4) What happens if the customer would like (or needs) to change the system provider? Is there any chance for him not to lose all his investments in time and money? - Nobody will take the risk of high or even total dependency on only one supplier (in the world) for such a function.

I am convinced that MT suppliers nowadays have learned their lessons, at least in theory. Therefore, the further development in "MT" systems will concentrate on three basic components:

- (1) Development of a translation management system (TMS) which includes all the functions needed to adapt the MT "function" to the user's domain and (textual) environment, like
  - filters to different (common) word processing systems,
  - text alignment and translation memory,
  - functions for a simple update of glossaries and MT dictionaries,

- pattern matching in pre- and post-processing.
- (2) Improvement of the MT function "as such", especially by providing:
- large (domain-specific) MT dictionaries,
  - context based semantic disambiguation (including sentence-overlapping pronominal reference).
- (3) Provision of tools which allow the customer at least to extract the lexical data and the TM material for "outside" applications. Development of "open" systems (see, for example, OPTIMIZER and TRADOS) which allow the integration of different MT tools.

### **Informative translation**

With informative translation I understand translation for users who don't want or are not able to ask for human translation.

There are many situations where this variant can occur. Without claiming completeness, some cases are cited:

- after accessing a data base, the result (an abstract, patent claim, ...) should be available in the user's mother tongue,
- a business letter has to be translated for better understanding the content,
- a newspaper or journal article has to be available in a language the user commands.

If we consider the increasing use of international telecommunicating networks like Internet, and even if we assume that English will dominate the international information traffic, it is evident that there will be a great need for informative translation, especially from English to other languages and from other languages to English.

In general, also for this market segment, we have to admit: the better the quality of the MT "kernel", the higher will be the consumer acceptance. Therefore, most of the components dealt with in the last chapter are also valid for this segment. But there are some other functions needed in this application environment:

- (1) There is no chance to "adapt" the customer to the system (with the exception that he must tell the system at least the target language for the translation). Therefore, the system must be able to identify - by analyzing the text and/or the "situation" - the relevant domain and even text type.
- (2) The system has to be very "transparent" to the user, especially by giving also information about possible problems of the resulting solution (this is, normally, also an aspect of machine-aided translation, but one could assume that human translators are able to find translation mistakes when comparing the original with the translation, which is not always the case in informative translation where at least one possibility is that the user has no knowledge of the source or target language).

- (3) The MT system has to be very "robust" in the sense that spelling or grammatical mistakes should not cause - at least in simpler cases - to translation failure or "crazy" word-by-word-translations. The "best" solution will be a grammar checker (and automatic corrector).
- (4) The system also needs a kind of "proper name" and "abbreviation" checker. In my opinion, there is no alternative to such an "automatic" function because this type of customer will not accept some pre-editing (like "blocking" of proper names from being formally translated ...).
- (5) Also in this application, the availability of large domain-specific dictionaries plays an important role.

It is quite clear that, in general, all the functions mentioned in both sub-areas of application, together, will contribute to higher acceptance of MT systems.

This is one - but not the only - requirement to be successful on the market. Therefore, it denotes the direction of investment and development. The next chapter will describe possible ways of distribution.

### **3. Distribution possibilities, MT services**

In the following, some main distribution and application possibilities are discussed:

- licensing and installation of in-house translation management systems;
- selling "small" PC-based products;
- offering one (or several) "MT online" host services within telecommunication networks on a purely technical level;
- offering complete (online) "translation services" including human high-quality post-editing.

#### **Inhouse installation**

On several reasons, there will be a need for installation of "larger" applications at customers' locations. Such a system cannot be compared with the "small" PC products: All functions of a TMS will be available, a special customer support has to be installed.

At the moment, this way of distribution is one of the "common" solutions, but the potential clientele still is hesitating because the functionality mentioned above is not fully available.

#### **Small "PC" based translation software**

At the moment, there seems to be a market for selling such products. Especially small and medium companies and individuals which have some need in doing "personal" translations are a clientele for such types of products.

Because the interface is very easy-to-use (GUI under Windows), the handling is simple, especially compared to the (existing) MTS.

There is a great chance that such systems will contribute to both the professional or semi-professional translation and the informative translation component.

For distribution, several variants can be used: provision together with big software providers (Microsoft, IBM/Lotus, ...); distribution by trade channels, dissemination by telecommunication networks ...

### **Online MT services**

This service will reach customers with incidental need of translation including customers with rare demand of translation in a special language.

If the developer of the system itself provides this service, the "full" power of the system with all its functionality can be provided.

One of the possible variants of such an application could be the combination with other data base services.

### **Complete online translation services**

This service type combines the machine translation part with human translation (or post-editing). The solution is very close to the "inhouse" solution with the exception that "experts" of language pair and domain (maybe located throughout the world) will be available to provide high-quality translations.

## **4. Conclusions**

The aim of this presentation has been to focus the viewpoint of research and development in MT on the industrializing part of the problem.

Not all of the problems have been mentioned. Especially there is a difficulty of standardizing in language industry so that customers can rely on the products to a higher degree.

In my opinion, there is no need of "persuading" the deciders or managers at least in language industry that MT systems will meet the market needs. Today's examples are the proof.

On the other side, if such a parallel is admitted, compared to aviation or automotive industry, MT and language industry are not in the "nineties". There is still a lot of investment both on the financial and intellectual/technical side.

The "external" conditions for successful activities in this business are very promising (and well-known): smaller computers with higher speed, more customers working in international communication, improved telecommunication possibilities with lower prices.