

Evaluation of Machine Translation Systems from a User's Viewpoint Some critical comments

Ursula Bernhard

Gesellschaft für Mathematik und Datenverarbeitung mbH (GMD)

Abteilung für Informationswesen

Sankt Augustin

The Gesellschaft für Mathematik und Datenverarbeitung mbH (GMD) is the German national research centre for computer science. Though coming from such a highly specialised institute I am no computer scientist, but a translator and a user of machine translation systems.

I have been using the LOGOS machine translation system since 1986. The application is successful as I have doubled the throughput of the translation service. Acceptance problems have not occurred since I am the only translator using the system and since it was my decision to have the system. In December 1989 I switched to METAL for reasons to be presented later. Though I am currently using METAL for my everyday work – 80 % of the texts to be translated from German into English are processed by means of METAL – LOGOS continues to be installed and I am very happy about that fact since it enables me to compare the two systems. In addition, there are translations LOGOS is more suitable for.

The evaluation of machine translation systems has been so far the matter of DP experts and computational linguists. For users, such evaluation activities are mostly a most academic thing since they hardly consider user requirements. Has the evaluation of machine translation systems ever taken account of the everyday work of the immediate users, namely the translators, a work which is characterized by pressure of time and low esteem? Has evaluation ever considered the fact that the potential users of machine translation systems are hardly trained for using technical means and that they are simply afraid of using computerized tools? Has anybody ever done a really practice-oriented cost-benefit analysis of the use of machine translation?

There is obviously a gap between theory and practice in machine translation and for bridging this gap, theory and practice should cooperate. This

is the aim of the present paper which will tell about the experience and the concern of a user.

The migration from LOGOS to METAL, for example, was not based on an evaluation of the two systems revealing the better quality of METAL, but on some most pragmatic considerations.

- An insufficient response to bug reports created the impression that a further development of the LOGOS software was doubtful or at least hardly promising. The METAL software seemed to be more advanced and promising, and the Munich-based development centre for German as source language made us hope for a better and quicker response.
- A considerable reduction of personpower at LOGOS Germany made us doubt whether LOGOS would survive. METAL was marketed as a Siemens product, and Siemens is one of the most powerful companies in Germany.
- At the time of decision-making LOGOS was marketed on a leasing basis while METAL was marketed on a purchasing basis. We needed to buy only the METAL software, the required hardware was available.

This last item was actually the decisive item due to the shortage of money in the research area.

Today I am rather happy that I have not tried to justify the migration by the better translation quality of METAL since I really do not know if METAL is of a better quality. Currently, I think, both systems are equally good or equally bad.

From a user's point of view the available machine translation systems show that they have been developed far away from practice. Let me illustrate this by some examples from my experience with LOGOS and METAL.

A comfortable user interface is most important for a software to be used by people who are no DP experts. This also applies to machine translation systems which are to be used by translators, technical writers or even secretaries as sometimes advertised. Therefore I think that even a linguistically sophisticated system translating the most complicated sentences correctly and resolving all ambiguities successfully will be of little use if its interface is uncomfortable and does not provide a smooth and quick handling by somebody who is normally not familiar with computer systems.

LOGOS has a homogeneous user interface since it runs on one computer system. METAL runs on two computer systems. The user has to struggle with two different keyboards, two different menu organizations and, in particular, with two totally different philosophies underlying the systems. Is such an interface user-friendly?

Machine translation systems are very expensive and much money has to be invested in hardware and software, training of personnel and organization.

Therefore, these systems will only pay if they increase productivity which means that all processes involved in translation have to be accelerated.

One of these processes is dictionary coding which is absolutely necessary to tailor the system to user requirements. By dictionary coding I understand the addition of new words to the dictionary by means of a menu-driven software component. The LOGOS dictionary coding is rather easy and fast. LOGOS requires information which is mainly restricted to the source language. Dictionary coding with METAL is more complicated. It requires a sound knowledge of both the source language and the target language since many grammatical details have to be entered. Coding takes more time since three entries have to be created, two monolingual entries and one transfer entry. LOGOS needs only one entry. This more detailed dictionary coding of METAL promises however a more controllable translation result while LOGOS sometimes behaves like a human by translating identical patterns differently.

Another problem of machine translation is pre-editing and post-editing. Pre-editing is often required for protecting words from translation, e.g. proper names or other untranslatable text material. In German software manuals, for example, it is quite usual to use a lot of English words which should of course not be translated. Therefore, it is sometimes most troublesome to protect untranslatable text material. LOGOS is tolerant towards unknown words in the input text which means that translation without the protection of untranslatable text material and even without the coding of unknown words is possible and delivers a result which is suitable as a basis for post-editing. METAL delivers a so-called phrasal analysis in the case of unknown words in the input text which means that a considerable amount of words is not translated at all. As for post-editing, as a translator, I hope that post-editing will continue to be necessary, otherwise the systems will actually become job killers.

Altogether, I think that machine translation systems should adapt to user requirements and not vice versa. METAL, for example, will get into trouble if a German sentence is longer than 30 words. In the worst case, the sentence is not translated, the system reports: error in translation. I think this shows clearly that the METAL development has not been very practice-oriented since German sentences in technical and scientific texts are often longer than 30 words.

These are only some first and most superficial impressions gained by a user using two different machine translation systems, but I think that such superficial or pragmatic factors should not be neglected when evaluating machine translation systems since they are important to the acceptability of these systems.