

DOROTHY SENEZ

The Machine Translation Help Desk and the Post-Editing Service

The growth of machine translation

Machine translation (MT), introduced to the European Commission in 1976, is one of the tools the Commission has developed to help its administrators cope with the pressures of working in a multilingual environment, and to ease the heavy workload of its translators. In its early years its use was restricted to the small number of individuals who had access to the Amdahl mainframe computer in Luxembourg. Then in the late eighties, the "raw machine translation" service was opened up on a help-yourself basis to all Commission staff, on both sides of the Ardennes, via the nascent electronic mail system. However, the limited availability of texts in electronic form, lack of information, the relative instability of the e-mail system, and the general unawareness of the very existence of MT all combined to keep its usage low.

In the early nineties electronic mail was coming to be accepted as a convenient and reliable method of interdepartmental communication. This, and a vigorous promotion campaign, created fertile ground for a dramatic growth in the use of machine translation usage at the European Commission. The number of processed pages totalled 30 000 in 1990 and 180 000 in 1995. The figure for 1997 (200 000 pages) is likely to increase this year with the introduction of a new web interface which will make machine translation available to all the institutions of the European Union.

Machine Translation Help Desk

Now that machine translation is freely accessible to such a large population, the Machine Translation Help Desk has assumed an important role as the supplier of a number of back-up services. It provides technical assistance to users, answers their queries, offers solutions to the inevitable problems created by format conversions, and prepares source texts prior to their submission to the MT system. It also acts as an essential link between

the users and those who develop the technical and the linguistic aspects of the machine translation process. In regard to information technology, the Help Desk acts as a troubleshooter: it monitors the problems encountered by users, and ensures that their needs and difficulties have the constant attention of those who develop the system. On the linguistic side, the Help Desk channels feedback from users to the linguists working on the system's dictionaries and programs. The expertise acquired in all these areas makes it an ideal test-bed for new developments. Last, but by no means least, the Help Desk runs the post-editing service.

Origins of the Post-editing service (Post-édition rapide or PER)

The striking increase in machine translation usage by the institution's operating departments at the beginning of the decade pointed to a latent demand for urgent translations, which was not being satisfied by traditional translation channels. Raw MT was usually corrected by the users themselves, with the help of colleagues of the appropriate mother tongue. In some cases MT output was being distributed in raw form, leading the Translation Service to look for ways to rationalise the use of MT. The post-editing service saw the light of day when it was decided to offer MT users a service they could call on for a rapid revision of their machine translation output.

It became clear that those who used MT the most did so because they needed help with urgent texts and were greatly pressed for time. Hence, a post-editing service, if it was to be useful to them, had to work within strict deadlines. Tightly run, "paperless" procedures were introduced, based on electronic mail, and this was a progressive step. In the early days, however, unreliable electronic mail facilities meant that deadlines could be met only by hotfooting it round the different Commission buildings in Brussels delivering diskettes, somewhat breathlessly, by hand.

Expansion

From these very modest beginnings, the service began to expand more or less spontaneously as news spread informally between colleagues, mainly by word of mouth. The period 1993-1996 saw a regular annual increase in PER of 1 000 pages (from 3 000 to 5 000 pages). In 1997, total turnover was 9 000 pages, an increase of 4 000 during a year. This would indicate

that a real need is being met. There is good reason to believe that the service could expand considerably if a formal publicity campaign were undertaken. European Commission officials work in a multilingual environment and the need to access and distribute information in many languages other than their own is a day-to-day reality. High quality human translation may take longer than they can afford to wait. In many cases it is quite simply a question of having post-edited MT or no translation at all.

Organisational aspects of the post-editing service

The revision of the raw MT output is entrusted to a small network of external linguists who link with the Commission's PER service entirely by electronic mail. The majority of these post-editors are experienced translators who have learned post-editing skills in the course of the assignments they have carried out for the Commission. The end users of the service are administrators working in the various operating departments of the European Commission. They may telephone the post-editing service directly to ask for rapid revision of a raw text.

PER is aimed at Commission departments who have urgent translation needs. Deadlines are negotiated in a three-way process between the end-user, the PER service in the External Translation unit and the free-lance post-editor. They are met in as short a time as possible, generally 48 hours.

The languages most in demand from the post-editing service are combinations of English, French and German, the so-called "working languages" of the institution. Of the 16 language pairs offered by the Commission's MT system, 50% of PER requests concern combinations of French and English and 40% of requests concern German in relation to French or English. The remaining 10% is made up of Italian, Spanish and Portuguese in relation to English or French.

Levels of post-editing

The term "post-editing" is open to different interpretations. A professional translator may well take raw MT output as an initial draft which he then polishes up to such a high standard that the end user is unaware that MT has been involved in the process at all. This type of post-editing might be compared to "invisible mending". In this case the customer is only

interested in the final product, which he expects to be flawless. How this is achieved is of no concern to him.

PER, however, offers an alternative service based on *rapid* post-editing. Although the PER product is not accompanied by the seal of quality associated with traditional translation channels, it may be a perfectly viable option, provided that three conditions are met. In the first place, the customer needs the translated version of the text urgently. Secondly, the text is not destined for publication but will serve some temporary purpose. Finally, the customer makes an informed decision to opt for post-edited MT after weighing up the advantages of a faster service against the risk of loss of quality inherent in rapidly revised machine translation.

Which texts?

It follows that only certain texts are suitable for PER. Rapid revision of the output from the Commission's MT system is meant for urgent texts intended merely for information or for restricted circulation, such as working papers for internal meetings, minutes of meetings, technical report or annexes. Documents for publication, texts that are politically sensitive or that are destined to become legislation require top quality treatment. In other words, the PER service can offer the most appropriate solution for short-lived documents. Those who request translation must therefore be aware of the different categories of documents and the different options available to them. The PER service acts as a filter, redirecting customers where necessary to the more traditional channels. Ultimately, however, responsibility lies with the requester.

Quality assessment

As a matter of principle, then, although the PER service does carry out spot checks, quality assessment is entrusted to end-users themselves. This is a radically new approach. The standard practice is that it is the linguists, not the users, who exercise control of quality. Traditionally, translations do not leave the Translation Service until they meet with the strict criteria imposed. In the case of post-edited MT, however, responsibility for quality control shifts to the user himself. It is he, after all, who has opted for rapid post-edited quality rather than a highly polished translation, and it is he who is best able to judge if his needs have been adequately met. Generally

speaking, feedback from users reflects a high rate of satisfaction, and any negative evaluations are followed up. Sometimes an unfavourable verdict stems from the judgement of a third party, who was not involved in the initial decision to opt for a given level of MT output. Experience has shown that the acceptance of post-edited MT output hinges on the reader being fully aware that the text is essentially the product of a machine.

Attitude of the reader

Hence, documents processed by the PER service always bear the heading "Rapidly revised machine translation". Machine translation rarely breaks away from the linguistic patterns and stylistic features peculiar to the source language. This may make for an awkward style, but should not necessarily obscure the meaning. Duly forewarned, the reader makes a mental adjustment. He needs the information and knows that without the post-edited version it would be unavailable to him. Hence, he is prepared to overlook this strangeness in style, and concentrate on the message, in the same way that the foreign accent of a non-native speaker invites a more indulgent attitude on the part of the listener.

The challenge facing the post-editor

The quality of the raw machine translation depends on a number of factors inherent in the source text: style and syntax, accuracy of spelling and appropriate use of punctuation. Moreover, the quality of the language pairs varies considerably in the European Commission's MT system. Combinations of French and English will give more satisfactory results than those language pairs which include German. In short, the output from the MT system tends to be unpredictable and the amount of post-editing required may vary upwards or downwards. Post-editing techniques themselves vary considerably from one language pair to another. French machine-translated into English, for example, is generally full of redundancies. Merely deleting the superfluous "the" can go a long way to rendering English target text acceptable.

The experienced post-editor no longer feels threatened by the machine, but has learned to reap as much benefit as possible from what the computer gives him. Often it is just a case of rearranging what is there. Global replace functions and macros can speed up the process

considerably. The result may not be couched in exactly the same terms he would have used himself, but if it conveys the correct message he can resist the temptation to touch it. Only that which makes little or no sense, or is actually wrong needs to be changed.

The aim of post-editing is to save time, and the challenge facing the post-editor is to reproduce the meaning of the original text speedily with as few changes as possible to the raw MT output. A post-edited text should be a grammatically correct and faithful reproduction of the original message. There is a very real danger of providing the reader with a translation which is readable and comprehensible but not what the original means. Particular care has to be taken to ensure that any distortions of meaning introduced by the machine are removed. Examples of common pitfalls are the use of negatives and the correct expression of tenses. As mentioned earlier, some stylistic inelegance may be tolerated, provided that the target text conveys the original meaning. In other words, a post-edited text should be cohesive, readable, and above all reliable. As the main aim is to convey the information contained in the text, there is no need to vary expressions for the sake of elegance. The reader has asked for rapid post-editing and he is not expecting pristine prose.

Future prospects

A service offering post-editing of machine translation on this scale is a new, and pioneering venture at the European Commission. It can absorb a very specific and meaningful part of the demand for urgent translations within the institution. Its advantages are twofold: speed and the availability of otherwise unavailable information. It is useful in cases where it has not been possible to plan ahead and allow time for human translation. Hence, it is particularly suited to short-lived texts. The initial, experimental phase of PER has been marked by spontaneous growth, despite an almost total lack of publicity. The rate at which it continues to grow will depend to a large extent on the backing it receives and its ability to maintain its streamlined procedures.

DOROTHY SENEZ

Bibliography

- FRASER BILL (1997) "The New Rhetoric" 149/1 79 *Terminologie et Traduction* 3
- LOFFLER-LAURIAN ANNE-MARIE (1996) *La Traduction Automatique* Presses Universitaires du Septentrion
- SPALINK KARIN / LEVY RACH / MERRILL CARLA (1997) *The Level Edit, Post-editing Process, A tutorial for Post-Editors of Machine Translation Output* Internationalisation and Translation Services Inc
- WAGNER EMMA (1985) "Post-editing Systran - A challenge for Commission Translators" 1/8 *Terminologie et Traduction* 3
- WHEELER PETER (1995) *Internal Post-Editing Guidelines*, Antler Translation Services Inc