

MT AS A COMMERCIAL SERVICE: THREE CASE STUDIES

Terence Lewis

Hook & Hatton Ltd
34 Central Avenue
Whitehills
Northampton NN2 8DZ
United Kingdom
E-mail: Hook_Hatton@compuserve.com

Abstract

This paper presents three cases studies showing the considerably different uses customers make of our Dutch-English MT service.

Introduction

Our company was established in 1985 as a technical translation company. In the earlier part of this decade it developed its own Dutch-English machine translation application with the aim of reducing the cost of translating large repetitive technical documents. When one of the key players in the business became seriously handicapped as a result of a stroke, a decision was made to make the MT program our unique selling point and operate primarily as an MT service provider.

The company's MT system is a modular, rule-based application which has been designed in such a way that it is relatively easy to adapt rules and write new rules "on the fly". Although the program uses many of the MT techniques described in the literature, machine translation has been tackled primarily as a "computer problem". Our approach is empirical in the sense that the program is continually refined and expanded in response to specific problems identified by "real" translation output. The application has been designed for use in a "practitioner-as-developer" environment; neither general nor limited licensing of the software has ever been contemplated.

The company markets its MT services through mailshots to Dutch and Belgian engineering companies. We offer a thousand-word free sample, and usually spend up to one hour doing terminology work for each sample, the aim being to demonstrate the program operating at full potential. We retail MT at a flat rate of 100 NLG (approx. \$53) per thousand words. Output is sold unrevised and with a disclaimer of liability. As a general rule, however, the target text is scanned for the results of "runtime errors" or for "howlers" produced by mistakes or omissions in the input file. A maximum of five minutes a page is usually spent on this "scan check". If it appears that longer would be required, additional dictionary work is done and the file is reprocessed. Source texts have to be supplied to us in electronic format via e-mail or on a floppy disk. An OCR service is available at an additional charge of around \$5 per page. This charge is increased if the text is faxed and may be reduced in the case of large-volume jobs. As the MT application can only handle text files, source format has to be manually overlaid onto the English output, and this work is billed at a variable hourly rate.

The quality of our MT output is directly related to the system's lexical coverage of the source text and the breadth of the subject matter. "Near human quality" has been attained with abstracts on artificial intelligence projects, cardiovascular reports, project specifications and simple system descriptions. Translations of documents involving two or three disparate disciplines are often problematic, and our poorest translations have been in subject areas with the least dictionary coverage. Dictionary building is still a key activity both at project level and as part of our ongoing development work.

After overcoming some initial skepticism the company is slowly building up a loyal customer base for unrevised MT. None of our customers sees our service as the sole way to meet their translation requirements. We are perceived, rather, as a specialist outfit offering a facility to which customers can turn to solve specific problems. Most of these are smaller companies which do not have in-house translation departments, although the Dutch subsidiaries of several multinationals come to us from time to time with rush jobs which nobody else would even consider. Occasionally we will receive an order from a Dutch employee who is required to write a report in English and wants us to provide something he can then “play around with”.

The following three examples highlight the variety of purposes for which industrial customers may purchase our unrevised MT output.

Case A: the multinational

Company A is the Dutch subsidiary of a huge multinational power and telecommunications group. It has an international staff and English is the corporate language. Many of its staff who do not read Dutch need to respond quickly to a variety of technical documents such as specifications issued in conjunction with calls for bids from Dutch government agencies. These documents generally total around 20,000 words and are required within 48 hours. We receive a fax informing us of the location of the source files and requesting a proposed delivery date. We download the files and confirm that we can handle the assignment. After processing the files, we log on to the customer's system and transfer the English output (using the same filename with the extension .ENG).

Despite the fact that this customer does not provide us with any terminology support, the MT output is found quite acceptable by the non-Dutch engineers who request the translations. This is partly due to our substantial dictionaries in the fields of computer engineering and telecommunications, particularly for system network architecture. In some files the only “unknown” words are proprietary names. The successful processing of these files is aided by the predictability of the style in which such documents tend to be written. Sentences are in the present or future tense, and rarely contain more than one subordinate clause. Constructions with the auxiliaries “kunnen” (can), “moeten” (must) and “dienen” (should) are very common, and the passive voice seems to be favored (“the system must be delivered...”, etc.). In our experience, this customer's documents contain between 20 and 30 key technical terms which are used over and over again. These terms are usually found in the Introduction, the Tables of Contents and the Summary (if there is one). If time allows we run these parts of the document first and check the dictionary coverage. We may also generate a list of all the words in the document and run that through the system. Failure to identify source words reduces the chance of a perfectly constructed output sentence.

Example A is an brief extract from a document recently processed for this customer.

Example A

Place of central equipment

The central equipment, such as the ISRA-cabinet (supply PTT Telecom), the CRE and the CAE, must be capable of being placed in the specially designed equipment room or computer centre. The tenderer indicates the conditions that are necessary for proper operation of the equipment such as environmental requirements and provisions to be made. Furthermore the tenderer indicates the physical scope of central and remote system(s).

The phrase “must be capable of being” is an all-purpose work-around translation intended to handle most cases in which the common Dutch verb combination “moeten + kunnen + worden + past

Participle” (= must + can + be + past participle) is used. A translation such as “must be designed in such a way that it can be + past participle” would be more elegant here, but would not be the appropriate translation in every context. None of our MT customers has objected to such “off-the-shelf” translations or to other all-purpose phrase dictionary translations such as “provisions to be made” for “te treffen voorzieningen” (literally = “to strike provisions”). The main purpose of the documents sent to us by this customer is to specify the items for which the tenderer has to submit a bid, the standards with which his equipment must comply and the delivery times to be met. All this information can be successfully conveyed by our program, despite the clumsiness of the translations of some of the individual sentences.

Case B: the small business

Company B is a small Dutch firm that manufactures natural health products. It updates its product literature several times a year, and the new documents are usually slightly modified versions of previous documents. The company does not have its own translation staff and the managing director’s secretary seems to be the only person in the business with a reasonable knowledge of English. The primary reason for using our MT services is cost. We can deliver usable computer translations at roughly 25% of the cost of human translations. The MT output is sent on to the company’s English distributor who checks our documents based on his own product knowledge and produces the final versions. As this customer does not have e-mail facilities, source and target texts are sent by “snail mail” on floppy disks.

This company’s product range is made from the same basic ingredients, with a number of specific constituents being added to create individual products with a particular effect. About 60% of the content of its product data sheets is identical. These typically begin with an introduction to the product which highlights its benefits, and go on to describe the main ingredients. The last part of the document deals with the dosage and any side effects. The documents are written in very simple language. The company provided us with translations of the main ingredients at the beginning of our relationship but nobody is available to assist with “on-the-job” inquiries. The customer has specifically declined our offer to post-edit the texts for a small additional charge, and the major share of responsibility for getting everything right seems to be placed on the shoulders of the English distributor who presumably does the work for nothing. Example B contains a typical extract from the documents machine-translated for this customer.

Example B

Pumpkin seed oil is rich in zinc and phytosterols. Zinc is very important for the metabolism of the prostate and frequently causes a drastic improvement. The phytosterol curbs an enzyme in the prostate that converts the male sex hormone (testosterone) into a powerful hormone which causes the prostate to swell. The oil, partly because of containing vitamin E and the substance curcubitacin, improves at the same time the tonus of the bladder muscles and leads to a relaxation of the sphincter, so that passing water becomes noticeably easier.

All the medical and chemical terms are correctly translated. “Partly because of” is a stock translation taken from the phrase dictionary. “Drastic improvement” is an unfortunate literal translation of the Dutch, which the English distributor would obviously change. If we felt that this word combination were likely to recur we could enter a better translation in the phrase dictionary. Another clumsy translation, “curbs an enzyme”, could also be improved upon by adding an entry to the phrase or pattern dictionaries. None of these or similar examples of poor translation has ever been criticized by the customer. On very large projects, we run off a few test pieces and do some dictionary work before processing the job in chief. Since this particular customer’s orders rarely amount to more than 5,000

words, that approach would prove uneconomical on a routine basis. As this is a fairly restricted domain and we have worked for this customer for some years, we now find that practically all individual words are correctly translated in the computer-drafted translation.

Company B is very happy with our MT service. Despite the large gap between our typical output and the best human translation, this customer has never complained about the quality of the translations or asked if the computer could come up with something more idiomatic or natural-sounding.

Case C: the reluctant corporate translation department

Customer C is the well-staffed corporate translation department of a multinational giant in the electronics industry. The translation manager at this company had known about our MT services for several years but had made a decision to implement a translation memory solution and use outside freelance translators to handle any overflow. Recently, however, this translation department found itself tasked with the translation of legal and financial documents totaling several million words to be delivered in about four months. Clutching at straws, the translation manager was forced to turn to us. Once authorization was obtained to allow these confidential documents off-site, we agreed to make our MT facility available on a “spare capacity” basis so that we would not neglect any of our existing customers.

It became clear to us at an early stage that the staff translators were not happy about the introduction of our MT services. Our MT output was heavily criticized by the revisers, even though various sample translations had been approved by the corporate lawyers who were the end users of the translations. Admittedly, the quality of our output was poor when compared with our computer translations of technical texts. This was primarily due to a lack of legal and financial terminology. In particular, the Dutch names of the Belgian government departments and laws had never been entered in our database as compound nouns or noun phrases. Sentence analysis was hampered by the considerable differences between the sentence patterns found in legal texts and those taken from technical documents on which our system had been trained. The actual mix of documents which were from Dutch and Belgian sources and varied in size between half a page and three pages proved to be an additional troublesome factor.

In view of the severity of the criticisms, we invited the translation department to share its own terminology with us. We then went on to provide the company's terminologist with some basic tagging rules which would enable us to use her terminology in our program with very little additional work. This move was well received, and after about a week we were confident enough to give the customer access to some quite complicated tagging rules for noun phrases. This level of co-operation led to a noticeable improvement in the quality of the MT output. There were very few subsequent criticisms, possibly because the in-house translators now felt they had a stake in the process.

The text shown in example C typifies the documentation involved in this project. These three companies are typical of the businesses which have used our MT services. For the smaller firms, the relative cheapness of machine translation seems to be the primary reason for choosing us. This is borne out by the willingness of several of these customers to wait a couple of weeks for translations which the computer can generate in a few minutes. Speed of delivery is the predominant consideration for the large corporations. Precisely what happens to our computer translations after they leave us seems to differ with virtually every customer. At Company A our MT output is transferred straight to the end user's mailbox without any post-editing by the translation department. Company B has our computer translations checked and edited by their English distributor who does not speak any Dutch. By contrast, the Corporate Translation Department at Company C saw our computer translations as no more than starting-points - very rough drafts to save them some typing; its translators considered themselves to be the principal agents in the translation process.

Example C

Please find enclosed, as previously agreed, coherent overviews of the companies that belong (or belonged) to the Group.

We classify these overviews as follows:

- a) overview of the Group as at 31/12/1991 together with the most important auctions and sales of companies that took place and/or expired before 20/04/1992.**
- b) overview of the Group as at 30/06/1992 together with the most important liquidations and sales of companies that took place and/or expired from 01/04/1992.**
- c) prognosis of the structure of the Group to be achieved as at 31/03/1993.**

We have proposed to restructure the Group in such a way to that it still consists principally of active operating companies, that are held by one or two (movie rights) parent companies.

Before 31/12/1992 the capital structure of the operating companies and of the parent company will be adapted, starting from new activities and continuity principles.

The advantages of bespoke software

We have also purchased licenses to several low-cost commercial MT packages and offer a limited German-English MT service to some local customers in the UK. As a result of considerable dictionary work, we are able to generate “engineer usable” translations in some specialized fields. For purposes of comparison we have been using Transcend for one customer’s machinery manuals and Professional Translator Plus for another customer’s reports in the field of photogrammetry. Use of these packages, which are certainly superior to our own application in terms of GUI and general operating convenience, has led to an appreciation of the advantages of being “keepers of the code”. In our own application, we have devised a simple rule-writing language which allows us to generate new rules to deal with specific syntactic and semantic problems. We miss that facility and other features in both these low-cost commercial packages. We now believe that it would unwise to base an MT service on a commercial package unless the vendor is able to guarantee easy access to the developers and regular upgrades in response to feedback.

Conclusion

Based on our experience, we conclude that suppliers of unrevised MT have a multi-functional place in the language services market. The purposes and uses for which machine translation output is purchased are diverse. However, as the “mystique” surrounding MT wears off, users are likely to become more demanding in terms of the accuracy of terminology and the general readability of output.