

Translator's Workbenches: A Practical Application

Adriane Rinsche

Abstract

The paper presents an example of a solution designed for a multinational client involving translator's workbenches.

As a first stage, the user requirements were assessed, as concerns contents and form of documents, the technological base, human resources and volume and cost of translation.

The range of possible solutions was then evaluated, in terms of budget, technological base again, language combinations, user friendliness, customisability and linguistic features.

Recommendations were then made in terms of either setting up an in-house language business unit or an outsourcing solution. The client opted for an outsourcing solution. We converted, aligned and imported existing information into the workbench solution and resolved organisational issues.

We have offered MT as an integral part of Workbench solutions to a variety of clients. We will discuss the financial viability and our own views of integrating MT into a translation technology environment.

Dr. Adriane Rinsche

Dr. Adriane Rinsche is currently managing director of The Language Technology Centre Ltd. She studied English and Philosophy at a German University, was a Research Assistant for three years, has a PhD in Computational Linguistics, and worked as a language and language technology consultant with major multinational clients. She has more than 10 years experience in the language and IT industries.

The Language Technology Centre Ltd.

LTC specialises in:

- evaluation of major language technology developments
- software consultancy in multilingual documentation and computer-assisted language training
- delivering and implementing software solutions
- coding and converting language resources
- providing and managing outsourcing solutions in multilingual documentation

Dr. Adriane Rinsche

The Language Technology Centre Ltd.

27 Cotswold Close, GB-Kingston, Surrey KT2 7JN

Tel: +44 181 549 2359, Fax: +44 181 974 6994

E-mail: xe_s412@kingston.ac.uk, or 101365.676@compuserve.com

www: http://ourworld.compuserve.com/homepages/Language_Technology_Centre

Introduction

Translation technology offers solutions to a variety of multilingual documentation problems which may vary considerably from one prospective user or user organisation to the other.

The increasing number of tools and systems becoming available on the market place requires a careful evaluation as a basis for any far-reaching decision, because the investment envisaged is not restricted to the cost of the software solution chosen. In order to provide the best result, any off the shelf product or product configuration requires considerable customisation. The time and cost involved in designing and implementing an appropriate solution should not be underestimated. Very well defined and well organised procedures need to be introduced in order to achieve a cost-effective, high quality and high speed solution to the translation problem. Today, many international corporations do not organise the processes involved in the most appropriate way yet.

An example of a solution we designed for one of our multinational clients will be discussed below. It illustrates the most crucial steps, embracing

- an analysis of the user requirements
- an evaluation of the relevant technology
- recommendations and an implementation plan.

As our evaluation was based on the requirements of a specific client, it cannot be generalised. Although we developed global evaluation criteria for translation technology, which were, for instance, applied as a basis of our recent "OVUM Evaluates" report, the choice of a particular solution depends on the requirements of a given company. There is no global "best solution" or "best product" on the market.

Towards integrated multilingual document processing

Our client is an American multinational manufacturer of machine tools. The European Headquarters are based in the UK. The management within one department of this organisation asked us to provide a technology based solution to their translation requirements. We gathered the following information as a basis for further research and our final recommendations:

Step 1: Evaluating User Requirements (May 1995)

1. Contents and form of documents

1.1. Text types:

The company handles, in principle, four text types within the technical/engineering domain, three of them highly repetitive, the last has a more informal and chatty style.

- Repair Instructions, repetitive
- Consumer Manuals, repetitive
- Machinery manuals
- Training materials, non-repetitive

1.2. Information available in machine readable form:

A considerable amount of terminology (parts lists) and parallel text in seven languages is stored in machine readable form on a mainframe computer, all in capital letters. The information stored on the mainframe will not be supported beyond the end of 1995. It is therefore vital to retrieve the information and convert, store and reuse it in an appropriate environment.

1.3. Languages:

English is the source language, at the time of the evaluation 6 target languages need to be covered, to be extended to 12 target languages later.

1.4. Formatting and tagging:

Formatting and tagging requirements are practically non-existent in the text type "Repair Instruction", with the exception of occasional tables and statistics. Machinery and consumer manuals, on the other hand, are produced in Pagemaker.

2. The technology base

Several departments use Ami Pro as their word processing environment throughout, on networked PCs and workstations.

The graphics department uses one Macintosh powerstation and a number of smaller MACs with Pagemaker as their major input and output medium for linguistic data and QuarkXpress as their major medium for pictorial data. Some employees use other packages such as Word for Windows.

3. Human resources

A group of technical authors, illustrators, project managers is responsible for documentation.

Translations are carried out:

- partially in-house by trainees or technical staff with language skills
- partially by various translation agencies
- partially by low level bilingual staff in the various national markets.

This is a fragmented and unorganised way of organising multilingual documentation, with a high degree of inconsistency, at low speed, resulting in low quality output.

4. Volume and cost of translation

The volume of text material translated in the past in one department was comparatively low. The source text material consisted of approximately 20 000 words a year and was translated into 6 target languages, with an overall annual cost of £14,000. If additional requirements are met without technical support the cost of translation will rise to £250,000 for traditional external translation services in one department alone (calculated on the basis of £120 per 1000 words translation).

Other departments within the organisation have similar translation requirements in the same subject domains.

A limited budget for introducing translation technology is available to the department with very urgent translation requirements, and although we would have preferred a top to bottom approach where we would have aimed to introduce an all-embracing solution, we decided to adopt an initial "pilot" approach in order to help our client with their immediate needs and in order to support their unusual interest and motivation to dedicate themselves to a technologically oriented approach to multilingual documentation. The department is obviously aware of the fact that a centralised and unified approach to multilingual document production and management is not only more cost-effective and consistent, but that it provides them with critical influence on the company's international corporate image.

Step 2: Evaluating the range of possible solutions (June 1995)

Based on the information obtained from the client, we used the following strategy for evaluating the translation technology products available on the market:

1. Budget

In calculating the cost of supplying and implementing a computer-assisted translation solution, the following costs had to be taken into account:

- a) Purchasing price for hardware and software
- b) Customisation cost (e.g. converting data)

We considered the following systems:

MT systems considered: METAL, LOGOS, DP/TRANSLATOR, SYSTRAN
all too expensive, therefore ruled out

Power Translator Professional
cheap, not ruled out for financial reasons

LMT
not ruled out for financial reasons

Translation Workbenches considered: IBM Translation Manager
Eurolang Optimizer
Trados Translation Workbench
Star Transit
XL8
investment in principle acceptable

2. Language combinations

The very advanced IBM terminology system Translexis is multilingual, but the translation system LMT with which it is integrated offers only two commercially available language pairs at present: English to German and German to English. The solution was not further evaluated for this reason.

The Power Translator Professional covers only 4 language pairs altogether, only 5 out of 8 language pairs offered by DP/Translator were relevant to the required solution.

Due to the high estimated degree of repetitiveness of 60% applicable to two out of three text types covering approximately two thirds of the documents involved, it was therefore appropriate to consider a translation memory based solution. All translation workbenches available on the marketplace offered the required language combinations, except the Eurolang Optimizer, which does not cover Greek. We retained the other four systems for further evaluation.

3. Formatting requirements

At the time of the evaluation, it was impossible to resolve all issues regarding layout and formatting, which ideally should be retained throughout the multilingual production process. Pagemaker and QuarkXpress were not supported by any of the translation products. The client requires further consultancy at a later stage in order to adopt an integrated and considerably more cost-effective strategy at a later stage.

4. Overall design

We distinguished between two categories of Workbench solutions:

Type A systems: database structure

Type B systems: flat file - reference material approach

Type A products: IBM Translation Manager,
Trados Workbench,

Type B products: Star Transit, XL8

We considered the Type A products more suitable in this context, the Type B products were therefore ruled out.

5. User-friendliness

Our client employs no linguistic staff in-house and does not intend to do so in the future. For the application planned it was therefore not feasible for the corporation to employ translators. It was our intention, however, to make the documentation process as transparent as possible to the client. We wanted them to be able to control the process from their premises to a certain extent by keeping one copy of the software on site. In order to do so, the system introduced had to be particularly user-friendly, with little training required, an adequate help system and a practical user-interface. Some features of the two systems which “survived” the evaluation to this final stage were:

IBM Translation Manager: System specific editor (at the time of the evaluation)

Project oriented design

Automatic translation memory update within project file

Maximum of 3 fuzzy matches displayed

List of not found words

no terminology database, flat file SGML dictionaries
instead

Trados Workbench: Integrated with Word 6
Integrated working environment for terminology and translation memory databases
Fuzzy matches up to user definable matching percentage
Access to bilingual concordance

6. Customisability

In order to meet the requirements of the specific application, a product chosen should allow for maximum customisability. Some features of the remaining products were:

IBM Translation Manager: Many WP and DTP systems supported

Flexible creation of dictionaries

User definable number and sequence of up to 10 dictionaries

alignment tool to build translation memories from parallel text

Trados Workbench:

Central translation memory database, user definable structure

(system, text and attribute fields)

User definable range of translation memory options

User definable automatic substitution of numbers, dates, times, names

Alignment tool to build translation memories from parallel text

7. Linguistic features

Translation memory should have identical and fuzzy matching capability, existing terminology and translated material should be easy to import.

Both solutions (Trados and IBM) had fuzzy matching capability and routines for morphological reduction. The Trados Workbench had the following additional features:

Trados Workbench:

Concept oriented terminology database design

Fuzzy matching at terminological level

Bilingual concordance

Neural network design (not a linguistic feature, but supports linguistic processing)

8. Conclusion

Our decision in this specific context was in favour of the Trados Translation Workbench. We expected a cost reduction for translation of approximately 60% for the repetitive text material, with considerably increased quality, consistency and

speed. The client requires further consultancy in terms of designing source document production with a view to multilingual documentation.

Step 3: Recommendations and implementation plan (August 1995)

1. Because of the highly repetitive structure of three out of four text types we recommended the above translation memory based approach with a terminology database both for terminology coding and update processing.
2. One multilingual terminology database and one copy of all bilingual translation memory databases required are held on site by our client. The reasons for this are several:
 - a) Data security is increased and risk management improved by having the data on more than one site. Although LTC holds copies of the databases on two powerful PCs and on tape, fire or burglary could result in the loss of the databases.
 - b) Any new text is submitted to the Workbench at the client's site to estimate the cost of a given translation cycle beforehand.
 - c) The percentage of identical matches is estimated to be 60%. The client is, in principle, able to increase this percentage by modifying very similar matches, using the database contents as their standard, and eliminating stylistic, syntactic and terminological variants. By this means an even higher percentage of the new source text becomes identical with source text already stored. This leads to a highly effective and standardised use of both language and storage capacities. In practise, with one year operational experience (August 1996) the client has relied on us for any source text standardisation.
3. The structure of the multilingual database and the translation memories is designed in such a way that a more comprehensive and integrated approach to language engineering within the overall corporate business processes can be implemented step by step.
4. LTC converts the contents of the existing mainframe database into Workbench format which provides instant reusable text and terminology in 7 languages and adds the relevant information for the additional 5 target languages in a second phase.
5. LTC manages the translation process on behalf of the client, as in-house linguistic staff is not available. LTC trains specialist translators in the use of the workbench and sends one fully updated copy of the databases to the client after each translation cycle. The client pays only for the translation of new text.
6. During a later stage in 1997 source document control will be increased by providing an English to English translation memory database. Standard text is stored in the database, new segments are accepted only if they are not variants of existing translation units.

After one year...

the solution is fully operational.

Statistics show that our estimate of an average reduction of 60% of the volume of translation and cost is correct. The example below was taken from our August '96 repair instruction update, German. There are 49% of identical text (cf. 100% below).

Repetitions are repeated untranslated elements, 50% of repeated words are charged to the client.

```
Start Analyze: Mon Aug 19 17:01:08 1996
Translation Memory: C:\PROJECTS\B&D\GERMAN\EN_DE1.TMW
C:\PROJECTS\B&D\GERMAN\R_ALL_EN.RTF
Match Types  Segments      Words  Percent  Placeables
Repetitions      250      1,063     11         0
100%             631      4,536     49         0
95% - 99%        28        290        3         0
85% - 94%        31        140        1         0
75% - 84%        20        150        1         0
50% - 74%        46        247        2         0
No Match         334      2,678     33         0
Total            1,340     9,104    100         0
Chars/Wo         4.28
```

We have an open strategy in terms of read and write access to the translation memory database. All translation work is, at present, carried out on site, therefore each translator can take advantage of maximum information, resulting in high motivation. Each translator is responsible for the status of his/her database, which increases motivation further. We are in the fortunate position to be able to agree any deadlines several weeks beforehand with the client to ensure that the same specialised and trained translators are available for each translation cycle.

The possible role of MT

We believe that machine translation plays an overall important role in information scanning and multilingual information retrieval, and will be increasingly relevant in value added and customised internet and intranet applications. In the past, considerable success was achieved with Machine translation and controlled language in technical documentation (cf. Caterpillar and Rank Xerox, to name only two examples).

With the emergence of translation workbenches, machine translation loses relevance for repetitive documentation in limited domains where a considerable amount of 100% matches can be expected through translation memory. However, machine translation could support the initial process of populating a translation memory database where parallel text is not available. Once a workbench is operational, it seems more practical to give translators terminology and translation memory information rather than having them post-edit machine output.

If machine translation is integrated into a workbench solution, design, implementation and production procedures should be as follows:

The first step is obviously to create or update the terminology database. Then all terms from the term bank are copied and imported into machine translation system dictionaries, at the same time adding the required linguistic coding. For the above client, this would have been possible, but expensive. The cost would have been GBP 2 per term = GBP 4000 per 2000 terms = GBP 48 000 for 12 languages, and the benefit would have been limited and restricted to only some target languages. The

client is therefore at present not interested in machine translation integrated into the current solution.

After updating the system dictionaries, the relevant source text can be submitted to a batch translation routine, with all identical segments automatically replaced by translation memory database content and all non-100% matches are machine translated.

In an interactive revision process the translator checks and post-edits the flagged machine translated text material and enters the correct sentences into translation memory. We would recommend this strategy to companies with a comparatively low repetition rate in their documentation.

