

2000

Terminology work: Tools and processes that make a difference

Uwe Muegge, *Monterey Institute of International Studies*

Terminology Work: Tools and Processes that Make a Difference

By Uwe Muegge

Technical texts, i.e., technical literature proper (data sheets, user documentation, scientific publications, etc.), as well as the whole range of medical and legal texts, have one feature in common: Their authors make generous use of: a) words not in common usage, e.g., *dongle* (a computer hardware device that prevents unauthorized use of protected software); and/or b) words that are in common usage but have a slightly, or even totally, different meaning in the special language, e.g., *bug* (in the general sense, this means a small insect, but in the computer software field, this is a small defect in the code of a program). Therefore, translators require access to domain-specific terminology data if they are to accurately transfer the meaning of technical texts from one language to another.

Unfortunately, terminology often varies within a given industry, for example, Sun Microsystems uses other terms than Microsoft for identical software features. Nor is it uncommon, even inside a single organization, for different terms to be used across divisions/product lines. To make things worse, new terms are coined and existing ones given new meanings in lock step with evolving science, technology, and law.

Terminology Standardization

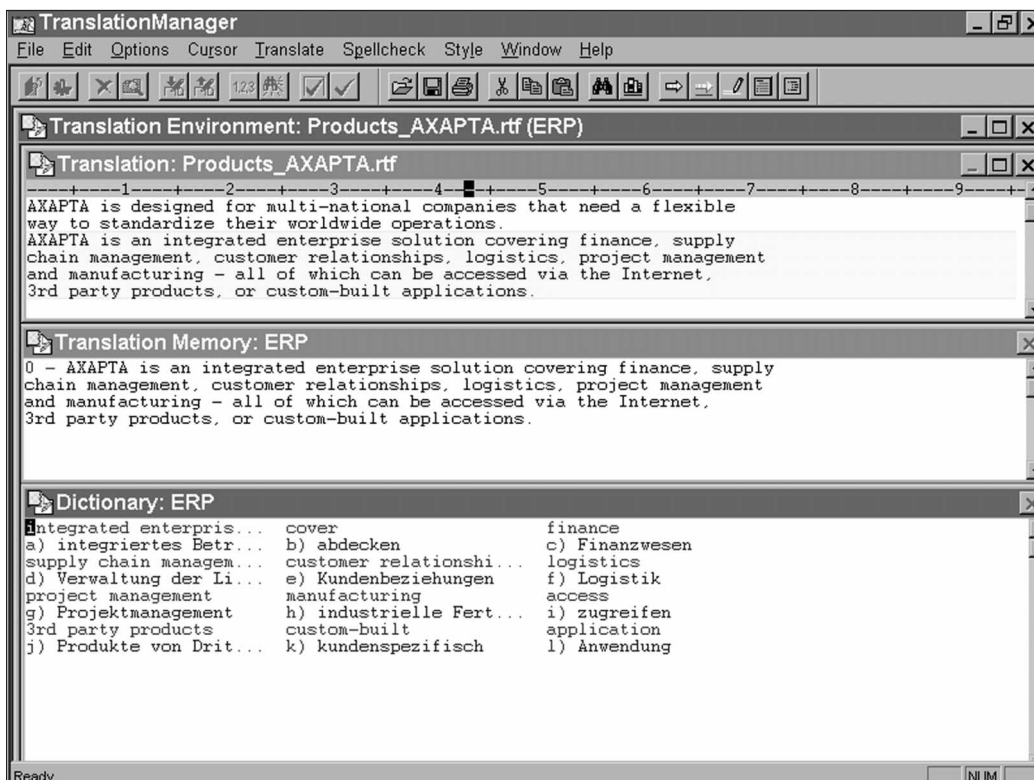
Yes, efforts are being made to standardize terminology. However, when I recently searched the Website of the International Organization for Standardization (<http://www.iso.ch>), the

...Analyze texts for unknown terminology as early as possible...

query “multilingual terminology” elicited a total of 63 hits. Considering the fact that many of these standards deal with terminology for terminology (e.g., ISO 704, ISO 860, ISO 1087, ISO 1951, ISO 10241, etc.), and that some sectors are covered by more than one

Continued on p. 16

Illustration 1: Translation memory system with automatic dictionary look-up



Terminology Work: Tools and Processes that Make a Difference Continued

standard (e.g., there are six for the petroleum industry alone), the somewhat sorry status quo is that standardized, multilingual terminologies are available for only a few dozen fields and span even fewer languages.

Conventional Resources for Terminology Work

In the absence of standardized, comprehensive, up-to-date terminologies for most fields and languages, the overwhelming majority of translators find it necessary to draw their terminological data from a wide variety of sources, for example:

- Glossaries supplied by the client or agency (the best-case scenario)
- Mono and bilingual dictionaries (printed and CD-ROM)
- Searchable public terminology databases (e.g., EURODICAUTOM, <http://eurodic.ip.lu>)
- Commercial terminology services (e.g., TermWright, <http://www.termwright.com>)
- Parallel literature (printed and Web documents)

In more than 10 years of industry experience, I have not once received a sizeable translation job with a *complete* list of all the special terms used therein. In fact, I have yet to meet the translation professional who receives *complete* glossaries with each and every translation assignment he or she is commissioned with.

It is safe to say that most technical translation assignments involve some kind of terminology work. Accordingly, you are probably only too familiar with the following problem.

The Quality/Cost/Time Dilemma

Most translation assignments are awarded on the basis of two limiting factors: cost (e.g., number of words times rate per word) and time (e.g., number of

days until delivery). If the text you are translating features a lot of unknown terminology (terms neither supplied by your client nor listed in your own terminology management system), you will have to put considerable effort into terminology work (contact client/experts, do research, build a terminology database, etc.).

Terminology work is always time-consuming: In the worst-case scenario, resolving terminology issues can consume more time than doing the actual translation. So if a translation project involves above-average amounts of terminology work to produce a quality product, the translation process will take longer than your average job with the same word count, the result being that your hourly earnings are lower than those from a “standard” job.

Three Golden Rules

Here is what I suggest in order to avoid situations where the amount of terminology work involved in a translation assignment jeopardizes your ability to finish the job on time and on budget—without exploiting yourself in the process:

1. Analyze the texts to be translated for unknown terminology as early in the translation process as possible—ideally before you make a time-cost commitment, and certainly before you begin translating.
2. Adopt a carrot-and-stick approach to your clients on this issue, i.e., reward their support and, if possible, charge a fee for performing terminology services.
3. Use the most efficient tools/processes and tap the most reliable resources when performing terminology work.

The Case for Translation Memory Systems

To secure terminological consistency within and *across* documents, translators must use a terminology management system. To secure consistency plus efficiency, translators must use a translation memory system (TMS) equipped with an integrated terminology management system.

High-end translation memory systems, such as Translator’s Workbench (<http://www.trados.com/products/workbench>), not only enable the reuse of translations, but also offer an automatic dictionary look-up function in addition to other great time-savers, such as automatic conversion and substitution of dates, numbers, and measurements, and bilingual concordance (a feature described in greater detail later in this article).

The automatic dictionary look-up function does exactly what the name implies: It automatically shows transfers for all terms in a given segment that are listed in the selected dictionary, and lets you insert these transfers with a click of the mouse.

Using a stand-alone terminology management system, translators have to actively look up every term that might be in their dictionary, meaning that they have to press a key or two for each and every suspected term. And because this is a tiring task, it is very tempting not to look up terms the translator already has a standard transfer for. For instance, *Diskettenlaufwerk* for *diskette drive* would normally be a perfectly acceptable German transfer, but if the client has specified *Floppylaufwerk* in their glossary, translators using any other solution than this one are committing a serious translation error.

For maximum efficiency, by the time you start the actual translation, your dictionary must include every term used in a given text (see Illustration 1, page 15). You don't have to be a rocket engineer to figure out that it is quicker to research all unknown terms from A to Z prior to translation and then translate without interruption, rather than to work in a stop-and-go mode (translate, research terminology, translate, research terminology, *ad nauseam*).

The need to have a complete dictionary before the actual translation work begins is particularly pressing when a project is split up among a team of translators. Teams without instant access to a complete master dictionary are condemned to even lower productivity levels—by the additional management effort required to communicate newly established terminology to other team members, to settle differences in opinion, and to correct invalid transfers that may have been used in the meantime.

The bottom line is therefore: The use of a TMS makes a lot

of sense even in environments where there is little to no reuse of previous translations, the reason being that automatic dictionary look-up alone makes for a much more efficient translation process and ensures terminological consistency.

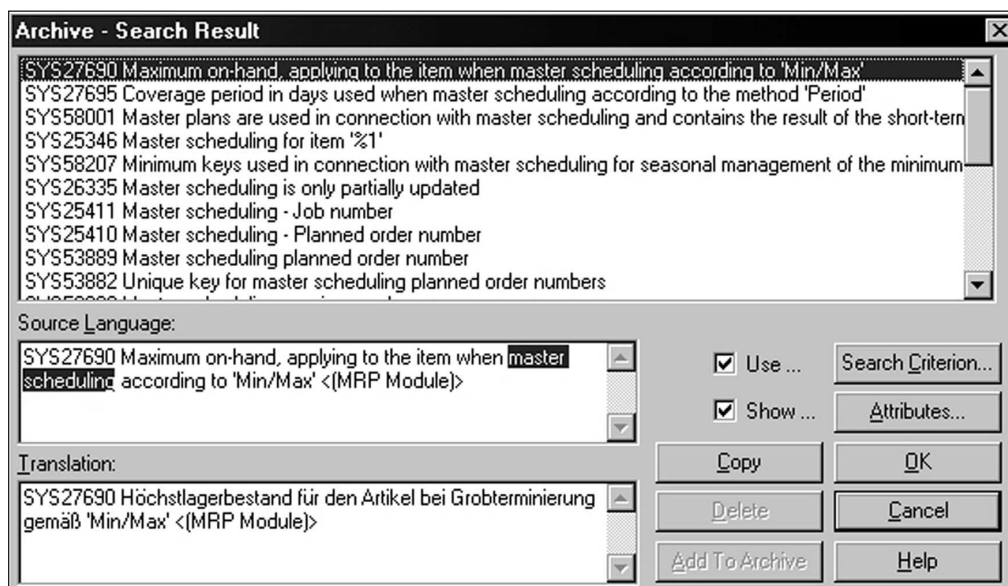
Research Translation Memories with a Concordance Tool

Fortunately, ever more translators are using a TMS. Even so, there is little evidence that building and/or maintaining a structured terminology database is part and parcel of every translation job. But if and when the call to build a project-, company-, or domain-specific terminology comes, large translation memories built by capable translators come in extremely handy. The terminology is all there, you just need to retrieve and sort it.

Many terminology management systems offer a concordance tool that can be used for this purpose. You

Continued on p. 18

Illustration 2: Concordance search in multiple translation memories using Personal Translator



Terminology Work: Tools and Processes that Make a Difference Continued

simply enter a term or sentence fragment and the system shows you all the segments in your translation memory containing your search item(s) or similar ones.

Unfortunately, with many leading TMS systems, including the Trados Workbench, concordance searches are limited by the fact that only one active and one look-up translation memory can be searched, and that translation memories need to be “inverted” in order to search segments in the target language.

So what do you do when you wish to research, say four different translation memories? For instance, when preparing a terminology database for a software product, you might use the translation memories of the string tables (lists of system messages), the help system, and the marketing material of the previous release in conjunction with the Microsoft Glossary turned into a translation memory for this purpose.

The tool I personally use for complex terminology projects is the “Search Archive” feature of Linguatrec’s Personal Translator 2000 Office Plus (<http://www.linguatrec.de/topics/mt2000.en.shtml>). With this concordance tool you can:

- Search up to 10 different translation memories simultaneously in a user-defined sequence;
- Enter up to four different search criteria, such as the beginning of a word, individual words, a string of words, or a combination thereof; and
- Search translated sentences without first inverting translation memories.

Best of all, you can use this tool from within Microsoft Word, i.e., in combination with, or instead of, Trados’ concordance function. You select an unknown term, press a shortcut key, identify the required term

in the target segment, copy all relevant information to your terminology database, and add whatever linguistic attributes your input model requires. Terminology research can’t get much easier than that.

Illustration 2 shows the results for a comprehensive search of translation memories built from string tables. Since in this project each translation memory covered a different product, individual translation memories were marked with user-defined attributes identifying the respective product (e.g., MRP Module).

At the time of writing, the Personal Translator 2000 only supports translation memories/archives in its proprietary format (i.e., users will have to write their own macros for converting third-party translation memories). For this reason, the powerful features Personal Translator offers are currently reserved for expert users. But if you don’t need to access more than a maximum of two translation memories at a time and don’t require sophisticated search functions to build your terminology database, the built-in capabilities of standard translation memory systems should work just fine for you.

Research Document Pairs with a Document Management System

So extracting terminology information from translation memories is easy—as long as the terms you are after have actually been used in the legacy translations your translation memories cover. But what if there are no translation memories available, only lots of source documents and their translations? Is turning them into a translation memory with an alignment tool like Trados’ WinAlign (<http://www.trados.com/products/winalign.htm>) really worth the effort?

Sentence alignment is certainly an option. The translation memory shown in Illustration 2, for example, was created specifically for the purpose of terminology research. In fact, aligning string tables is a piece of cake, since a numbered line in the source string table corresponds precisely to a numbered line in the target string table. Aligning this type of document doesn’t even require a sophisticated alignment tool—you can simply turn the two data sets into a two-column table and then convert it into your preferred translation memory format using a simple search-and-replace macro.

By the way, in this particular job, my low-tech alignment setup worked better than Trados’ WinAlign and IBM’s ITM, both of which kept crashing while crunching the 12,000 text lines each of these files contain.

So feed your files into an alignment tool and see how successful it is at automatic alignment. If there is not a lot of manual editing left to do, fine—go ahead and turn your file pairs into translation memories (which you then process as described in the preceding section). However, if aligning your

files promises to be time-consuming, just process your files in a document management system (DMS) instead.

A DMS, like PageKeeper Pro from Caere (<http://www.caere.com/products/pagekeeper>), allows you to index files of almost any type (e.g., printed documents, Web pages, word processor documents, etc.) and then use powerful search functions like Boolean operators, proximity searches, and fuzzy searches to quickly locate individual words, sentence fragments, etc., in and across documents.

The process is quite simple. In PageKeeper Pro, you add your source document to the DMS either by navigating to a file that's already on your hard drive or by importing a document that isn't. Again, this type of system handles printed documents (which you scan in and process with the built-in optical character recognition [OCR] facility) as well as all popular electronic file formats, including Web pages. Next, you add your target document or documents (if you have multiple documents, each can be in a different target language). Finally you link, or "clip," the source and target documents together. You can clip together documents regardless of their document type. For example, a Microsoft Word document can be clipped together with a scan of a printed brochure and a Web page.

It is important, however, that clipping is done in such a way as to maintain a high degree of parallelism between source and target documents. If you are dealing with a product catalog, for instance, in which each product group represents a chapter, and some product groups are not covered in the translated documents (perhaps because they are only offered on the domestic market), it is advisable to break up the catalog into chapters and clip together corresponding source and target chapters.

Now you are ready to search for terminology. Enter your term or phrase in the source language, click on the "Find" button, and you will almost instantly see a list of all documents containing the search term. When you open a document in the result list, the system takes you to the first occurrence of the

queried term (which will be highlighted in the document). Now you simply open the linked target document and go to the corresponding position in that document to find the transfer you are looking for.

For example, if the source term was found in the first sentence of the first paragraph of page 16, you would go to page 16 in the target document and examine the first sentence of the first paragraph. Depending on the degree of parallelism between your source and the target document, the transfer in the target language will be found either in that first sentence, or further down on page 16 (if the translation is longer than the original and the layout wasn't adjusted), or at the bottom of page 15 (if the target text is shorter than the original). Then you transfer the result of your query from the DMS to your terminology database using the copy-and-paste function.

Researching terminology in document pairs with a DMS such as Caere's PageKeeper Pro is almost as easy as searching translation memories with a concordance tool. You may just have to look a little longer for a transfer in the target language, as DMS systems link texts not at the sentence/segment level, but at the document level instead.

But then again, a DMS allows you to link one source document to any number of target documents. This means that researching terminology in more than one target language can be done in one go, whereas in a TMS you will have to open and search different translation memories for each language combination.

The bottom line is that a DMS is a great tool when it comes to performing terminology research on large pools of data such as legacy client information

Illustration 3: C-Pen offline scanner from C Technologies.
Photo © C Technologies



Continued on p. 20

Terminology Work: Tools and Processes that Make a Difference Continued

Illustration 4: Sample input sequence for creating a MultiTerm import file with an offline scanner

```
**
<German>Grobterminierung
<English>master scheduling
<POS>noun
<Definition>Master scheduling
analyzes your material and capacity
requirements here and now.
<Example>The master schedule has
the additional time fence options of
freeze period, scheduled release,
forecast plan, action message, and
future message.
<Source>AFS
**
```

and parallel literature—without having to first go through the painful process of sentence-by-sentence alignment.

Acquire Data from Printed Sources with an Offline Scanner

But what do you do if worst comes to worst—i.e., your client doesn't provide any useful material for terminology research; you don't have any relevant reference material; there is nothing on the Web; your subject-matter expert friend is on vacation; and the urgent requests for help you posted in every relevant newsgroup and forum yield exactly zilch?

Before throwing in the towel, check the catalogs of the libraries in your area first. If something turns up that might help, grab your list of unknown terms, your offline scanner, and hit the road.

An offline scanner, like the C-Pen from C Technologies (<http://www.cpen.com>), is basically a digital camera with built-in OCR functionality, which makes it an excellent input device for translators/terminologists who need to turn printed text into electronic data away from their office. Offline scanners are small enough to fit into your shirt pocket, can store hundreds to thousands of pages of text (depending on the model),

and require only two AAA-size batteries for weeks of service.

Terminology research involving printed documents at public libraries or other such offsite locations cannot be quite as efficient as the strategies described above involving electronic documents, because the search process proper must still be carried out by the translators/terminologists themselves. It is only at the data acquisition stage that technology can kick in.

And here is how. Once you have found the right transfer in a printed reference work, you just whip out your offline scanner and scan in any useful information, e.g., term, definition, sample sentence, and so forth. At reading speeds of up to six inches per second, offline scanning easily outperforms manual text input with a keyboard. Best of all, you will still be moving your scanner gracefully from line to line long after the laptop-lugging competition feels the need to rest wrists or recharge batteries.

And that is only the half of it. If done properly, you can build a complete terminology database right in your offline scanner, so that the data you bring back to the office can be imported to your terminology management system without further user intervention.

For this to work, your information needs to be scanned in the right sequence as discrete units of information, employing the delimiters used in your specific terminology database definition. This is actually simpler than it sounds. See Illustration 4 above for an example of an entry for import into a Trados MultiTerm terminology database.

There are seven delimiters in the sample (**, <German>, <English>, <POS>, <Definition>, <Example>, <Source>), each of which must be scanned in before the corresponding unit of information. These delimiters can be printed either in your list of unknown terms or in the footer of your printout if you are bringing your source document along. That's all there is to it.

Using an offline scanner rather than a laptop/palmtop computer or handwritten notes to acquire terminology information from printed sources away from your office is not only faster than these other methods. If used in combination with a smart input model, it also eliminates the need for subsequent manual data input, post-editing, and data conversion. That's efficiency for you!

Conclusion

Terminology work is of immense practical value. If done at the right time (i.e., prior to translation) and with the right tools, it will not only improve the quality of your translation product, but greatly accelerate the translation process as a whole. And if my experience is anything to go by, once you work in a translation memory environment—which automatically supplies transfers for every term in every sentence of your source text—you won't ever want to go back. Not to mention your clients, who'll pick up soon enough on the enhanced terminological consistency/quality and faster turnaround that these tools and processes enable.

ata