



Christine Bucher

Christine Bucher is a senior project manager at Iverson Language Associates, Inc., and can be reached at christine@iversonlang.com

PREPARING FILES FOR TRANSLATION

CHRISTINE BUCHER

Although much has been written about how to effectively write content for translation, another aspect of writing for translation has not received quite as much press — best practices in usage of authoring tools. Just as employing principles of good writing in English documentation helps to allow the translation to be correspondingly well written, good habits of electronic document creation help ensure that the final formatted translated document accurately reflects the English. This article, a brief look at some technical issues of preparing files for efficient translation and the repercussions if this is not done, focuses primarily on FrameMaker.

Auto-generated files

Best practices in document preparation dictate that you re-generate auto-generated files before finalizing your documents to send to print. In reality, a writer will often make last-minute corrections or changes directly in auto-generated files, rather than risk the possibility of introducing larger errors by re-generating the entire files. This approach, while it has immediate expediency in the English-only world, can have significant ramifications if the document is handed off that way to the translation provider.

Auto-generated lists

There are two main types of generated files in Adobe FrameMaker. The first type is based on visible elements in the document. Examples of this type of auto-generated file include tables of contents (TOC) and lists of figures (LOF). The second type is created by manually inserting invisible markers in the body of a document. The primary example of this type of auto-generated file is the index.

One of the great benefits of modern-day text processors is the ability to auto-generate lists from elements within a document. For example, one can automatically generate a TOC from heading styles that

are assigned to headings within the document. FrameMaker can also generate a LOF from a style that the user applies to figure captions within a document. When one auto-generates a TOC, FrameMaker

and/or the user manually tweaks the TOC file in order to correct it, the same process will be required for foreign-language translations. If last-minute changes are made to the TOC and the corresponding change is not made to the actual heading where it appears in the body of the document, not only will one lose the change in the English file the next time he or she generates the TOC, but the change in the translated file will also be lost because the translated TOC is generated from the translated headings in the body of the document.

Of course, one can still create the TOC manually, but that route is becoming less common. When a TOC is created manually, it is much more difficult to keep consistency between the name of a section as it appears in the TOC and the name as it appears in the body of the document. Any inconsistencies in the English will be duplicated in the translation. When we're performing quality assurance checks on a manual that has been handled that way and we ask the translator why the title of a particular section in the manual doesn't match the title in the TOC, the answer is usually "That is how it is in the English!"

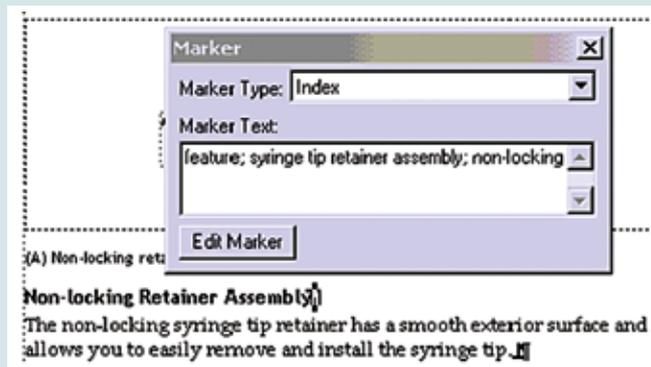


Figure 1: Index markers in English file.

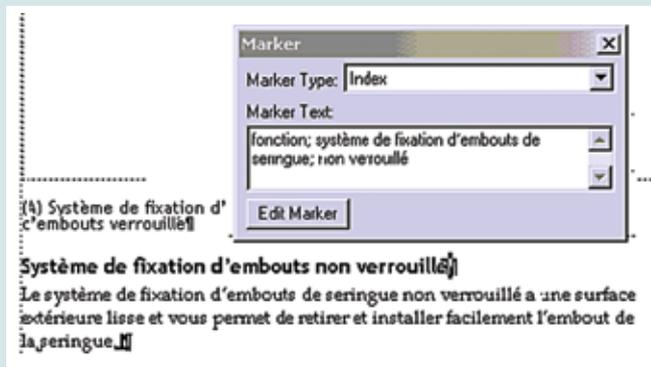


Figure 2: Index markers in French file.

creates a new file in the book and automatically appends TOC to the file name. When an LOF or list of tables (LOT) is auto-generated, FrameMaker creates the file in the book and automatically appends LOF or LOT respectively to the file name.

If one uses FrameMaker to automatically generate these kinds of front matter files for a document, the same process can be used when foreign language translations are formatted. If heading styles have been applied correctly and TOC styles assigned in the English manual, the process will be relatively seamless in the foreign-language manual. On the other hand, if the English heading styles have been applied incorrectly or incompletely in the English document

Indexes

The second type of auto-generated list in FrameMaker is created by manually inserting invisible markers in the body of a document. FrameMaker index files are generated from this type of marker. For example, the author inserts an index marker on a page that is to be referenced in the index and types the index entry text in a special index marker dialog box. The author then uses a separate process to generate the actual index page. FrameMaker automatically adds the new index file to the book and appends IX to the file name.

When a file containing index markers is prepared for translation using translation memory (TM) software such as SDLX, the

index marker text is extracted and put into a separate file. After translation, when the translation file is converted back to FrameMaker file format, SDLX inserts the foreign language into the marker text in the appropriate places in the foreign language FrameMaker manual. The foreign language index is then generated from the marker text.

Therefore, if one manually types an index entry and page number in an English index file, it will not show up in the foreign language index. Likewise, if one deletes an index entry in an English index file but does not delete the corresponding marker, it will auto-magically reappear in the foreign-language translation. Of course, a user who doesn't read the language won't know it is there!

Checking a foreign-language index against the English is inherently difficult. Foreign languages are sorted alphabetically according to the translated terms, not according to the English. In addition, there is not necessarily a one-to-one correspondence between an English entry and foreign language entry because a single English word may be translated in different ways in the foreign language and vice versa. So, if the user counts the number of index entries in the translation and compares it to the number of entries in the English, the numbers won't necessarily match. Failure to provide the translation provider with an up-to-date auto-generated index will compound the difficulty of checking index files.

Note that although European languages can take advantage of the automatic alphabetization capabilities of text processors such as FrameMaker or Word, this capability does not necessarily extend to languages that use non-roman fonts. Languages such as Chinese or Russian usually must be manually sorted by a translator. FrameMaker has the capability to specify sort order, for example, of Japanese kanji, but you have to know how to pronounce the kanji to do so.

Variables

Users of sophisticated features of text processing software such as FrameMaker's variables are aware of the benefits of consistency that variables bring to documents, as well as the time savings of not having to re-type long strings of repeated text. When documenting a new product whose name changes several times during the development cycle, by using variables one can quickly and consistently change the product names across several documents.

As beneficial as variables are in English, they can present problems on the translation

side due to different grammatical rules in other languages. For example, variable text that would have the same ending in English regardless of whether it is used as the subject or object in a sentence might have a different ending in another language depending on whether it is in the nominative or accusative case.

TM software extracts and re-inserts variable text from FrameMaker files in the same manner as it extracts and re-inserts index marker text. Variable text is first extracted into a separate file for translation. When the translated variables are re-inserted into the translated document, the exact same text string will be inserted, regardless of whether its placement in the sentence requires a change of case. Unfortunately, this is not something that can be remedied by making sure that one has properly updated auto-generated fields in the documents. But users should be aware when deciding to use variables in a document that the efficiency gained on the English side is counterbalanced by the complexity that may be added on the translation side. A linguistic review will be necessary after the text is converted back from the translation file format to the native file format to identify potential linguistic problems.

Hidden text

In a similar vein, one should be careful when using commonly used features of Microsoft Word, such as hidden fields or Tracked Changes — for example, applying the hidden text property for text that one wasn't sure he or she wanted to delete and then forgetting to delete it later or turning off "viewing Tracked Changes" but forgetting to remove the Tracked Changes before sending the file to your translation provider. Tracked Changes left in the document will usually be spotted immediately, but hidden text may be extracted along with the rest of the text into the file for translation. If the hidden text causes the sentence or fragment to be confusing, the translator may question it; otherwise, it might be translated along with the rest of the text. When the translation file format is converted back to Word, the part of the translation that corresponds to the hidden text will be hidden again in the translated file, but the remaining translation may be unclear or ungrammatical if it depends upon the hidden text for part of its meaning.

Hidden columns or rows in a Microsoft Excel spreadsheet present a similar situation. Although they may be easier to spot, one

should still communicate to the translation provider that the hidden text does not need to be translated, if that is the case.

Other considerations

A couple of other considerations in preparing electronic files for translation will make it easier for a translation provider to work with the documents.

Clean directories. Remove any old files, duplicate files and backup files from directories before sending files for translation. If one does not take the time to do this, the translation provider must pick through the files to determine what needs to be translated. Since the translation provider is not as familiar with the files as the client is, this process will not be as accurate as if the client had done it. The translation provider may only be able to identify obvious duplicates and unneeded files may be sent for translation, which the client may end up paying for unnecessarily.

File naming conventions. Create and adhere to standard naming conventions for documents. A short descriptive English name is best, but if numbering schemes are needed to identify documents, it helps a great deal if the scheme can be consistent. Think of how new writers on staff may struggle with cryptically named files until they internalize the file names. It is the same struggle for a translator who does not have the advantage of working closely with the client's team. If the translation project manager is handling a large volume of files into many languages, he or she can process documents with meaningful file names much more efficiently.

Summary

The best way to ensure consistency in English documents and streamline the translation portion of the process is to ensure that automatically generated files are re-generated immediately prior to finalizing the files and that no manual edits have been introduced in auto-generated files.

In the past, overburdened and time-crunched technical writing departments may have sometimes found it necessary to make manual edits to auto-generated files, knowing they could fix them with the next release. Now that translation has become part of the documentation cycle, it is important to hand off files that are as clean as possible to the translation provider. A better understanding of the impact downstream will, we hope, make the extra effort worthwhile. 