Localizing Indesign Documents

By Uwe Scholz and Roman Kotzsch of milengo Inc.

Shortly after the release of Adobe InDesign CS for Windows and Mac OS in 2003, it quickly became regarded as the new Desktop Publishing standard in document production. However, companies that publish InDesign documents in multiple languages still face challenges in developing appropriate procedures and automation tools for efficient and reliable localization processes. This article reviews four relevant conversion tools currently available on the market and highlights common problems in the localization process as well as how graphic design agencies, localization agencies and their clients can best avoid them.

Table of Contents

- Assumptions
- Tool Overview
- Pre- and Post-Processing
- Three Typical Issues
- Layout Guidelines for Preparing InDesign Documents
- Workflow Aspects and Quality Assurance
- Conclusion

Assumptions

This article is founded on the assumption that after the InDesign documents are created and approved by graphic design agencies or internal client DTP departments in one single language, companies normally use external vendors for translation (while document preparation is carried out in-house) or outsource the whole process to their localization provider.
The actual localization process that milengo follows typically includes the following steps:

- Preparation of source documents for export/import procedures (pre-processing)
- Export of translatable text from an InDesign file to a format that can be accessed by translation memory applications
- Translation of content by local milengo members
- Import of translated text into InDesign
- Adjustment of the document's layout and features (post-processing), and
- QA and approval.

There are three main types of features which are relevant for export/import before/after translation:

- **Text paragraphs** should still be in their original position (frame, page) after translation and the original formatting (paragraph and character styles) should be retained.

- **Image frames** are used in two ways in InDesign documents: at fixed positions related to page borders or within the text (floating). Both kinds of image frames should be retained and should stay in their original positions relative to the respective reference elements.
- The third type of feature is **index marks and hyperlinks within the text**, which are used to automatically create document indexes and (when exporting PDF documents) for hyperlinks to internal references, external documents or web sites. Marks and hyperlinks should be exported in a way that enables translators to translate the associated entry text along with the respective text phrases. After translation, all types of marks should be imported in a way that does not impair the internal processing of InDesign.

An ideal conversion tool would:
- Not require any pre- or post-processing of InDesign documents
- Export/import all relevant InDesign document features correctly
- Exports to a file format that is compatible with different translation memory tools
- Present translatable text to the translator in a layout that is as close as possible to the original InDesign document layout, especially regarding font formatting and tables.

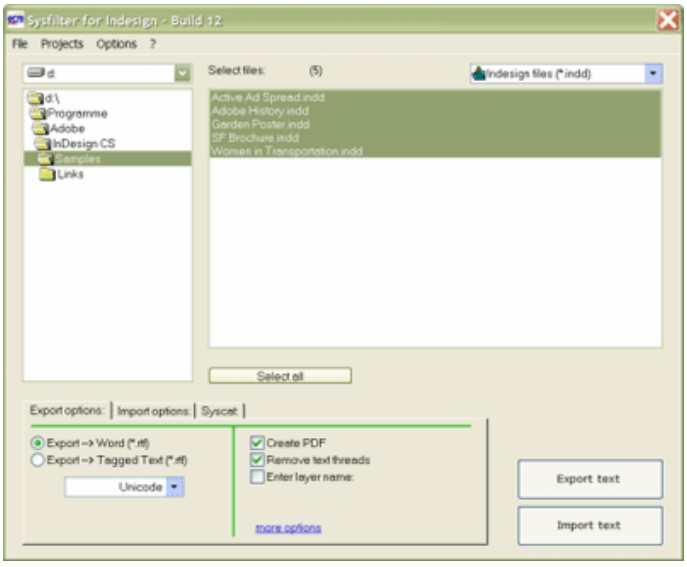
As the ideal conversion tool (still) does not exist, the localization agency will need to perform the pre- and post-processing. The actual work required for pre- and post-processing depends heavily on the respective conversion tool's ability to export and import document features. Therefore, we will first talk about the available conversion tools, and later consider some aspects on how to make the best possible use of them.

Tool Overview[1]

For the past decade, TRADOS S-Tagger has been the tool of choice for localizing FrameMaker documents. For InDesign files, no single conversion tool has been established yet as a comparable standard. The best known tools currently available for converting InDesign CS files into translatable file formats are TRADOS Story Collector for InDesign CS (released in 8/2004 with TRADOS LSP 6.5.5.438), SDLX InDesign CS Filter (released in 10/2004 with SDLX 2004 Bonus Release), STAR Transit InDesign CS Filter (first released in 5/2004 with Transit XV Service Pack 11) and ECM SysFilter for InDesign (first public version released in 3/2004). Table 1 compares the relevant features of these tools:

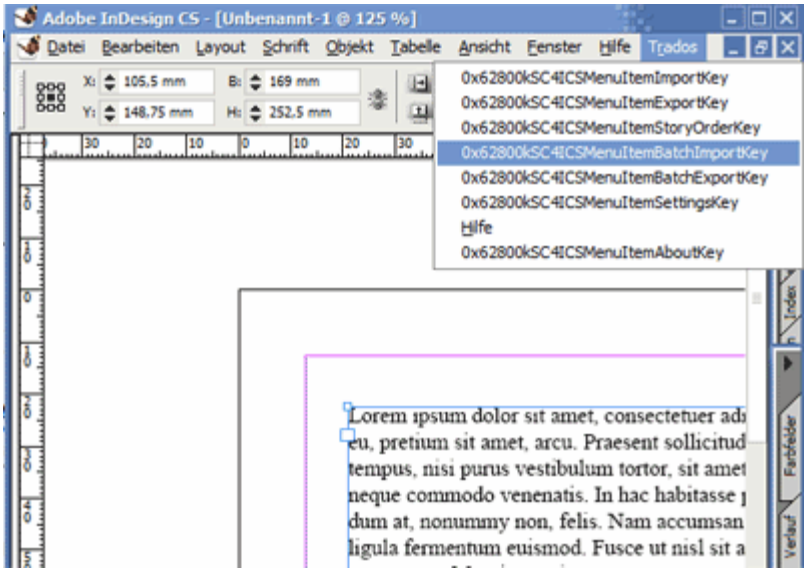
[1] milengo Inc. does not exclusively support or endorse any of the programs mentioned.

ECM SysFilter



ECM SysFilter exports translatable text into RTF formatted files. These files are especially appealing to translators because the original text layout, including tables, is transferred to the translatable file. Index marks can be edited directly in the text. However, if the InDesign document has inline images, a DTP engineer has to copy each image manually onto the InDesign pasteboard before exporting with the SysFilter. After translation, all images need to be manually pasted back into the text. This can be a time-consuming task when multiple target languages are involved. To make things worse, hyperlinks are stripped from the document during translation import, and need to be restored manually. On the other hand, the well-preserved document formatting and tables in the exported RTF file provide a valuable ”emergency option”: if the automatic import process fails, it is possible to copy and paste the translation manually from Microsoft Word to InDesign. The use of the standard RTF format gives LSPs and translators the freedom to choose their preferred CAT tools.

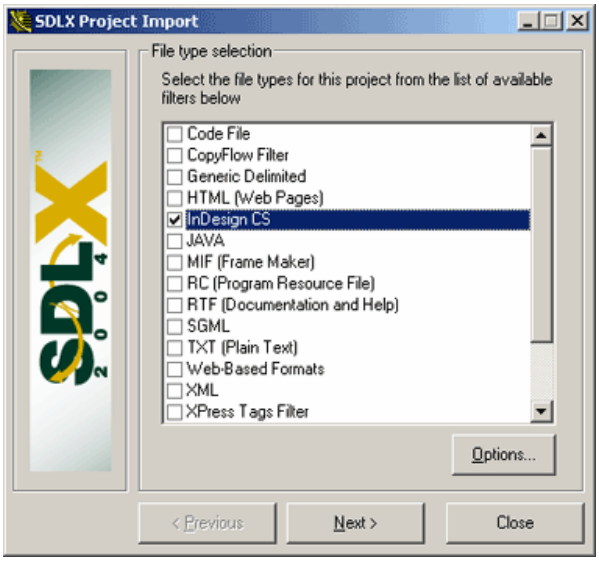
TRADOS Story Collector



While tables are difficult to recognize in the tagged format, tags are doubly protected: they are locked by default and TagEditor alerts translators about missing/wrong tags in fuzzy matched translations.

SDLX

SDLX exports translatable text into it's own ITD file format. By default, the table of contents (TOC) and index are ignored during export since they must be regenerated after translation anyhow. Inline images are handled well: they are placed correctly after importing the translated text into InDesign. Index marks can be translated in SDLX and translators can easily recognize them by noticing the square brackets. Their relation to text terms, however, is not always obvious. Hyperlinks are stripped from the InDesign document during translation import and need to be restored manually. SDLX recently released a small tool for Trados users designed to convert between ITD and TTX formats.



STAR Transit

The Transit Filter handles all relevant features correctly, and is therefore quite close to an 'ideal' tool. The Transit user interface helps translators to understand the main text structure by presenting bold text in bold face and table text in tables. In our tests, however, the formatting from InDesign paragraph styles was not shown in the translator's interface. Unfortunately, the exported TTC format is proprietary and files can only be translated with Transit.

Pre- and Post-Processing

Depending on the export tool used, hyperlinks, auto-generated text and inline images may need to be removed from text frames before the translatable text is exported, as described above. It might also be useful to adjust the document to some of the layout guidelines described below, which will in many cases simplify post-translation processing. Before beginning the translation, a test import of the still untranslated text should be performed for each converted document to confirm the accuracy of the export/import procedure. Since the Trados- and SDLX-exported text does not provide a reasonable overview of the original text structure, DTP Engineers should create a PDF file of the source document for the translators to use as a reference.

After importing the translated text, the target language attribute should be applied to all styles to ensure correct hyphenation. Next, all text elements need to be checked and adjusted for the complete text display, as well as text positioning and formatting. Hyperlinks and inline images must be restored. Finally, the table of contents and index need to be generated to finalize the process.

The manual work required to complete the full pre- and post-processing is time intensive, error-prone and more costly compared with the conversion of FrameMaker files for example. To better streamline the process, milengo decided to embark on the development of a proprietary InDesign plug-in that automates the pre- and post-processing tasks associated with InDesign files. Without any further

manual editing, inline frames, hyperlinks and bookmark sections can either be fully exported into RTF format, or the InDesign files cane be pre-processed to use with the preferred export CAT tool.

Three Typical Issues

Whatever tool you decide to use, InDesign itself might prove to be an issue. InDesign files from versions 2.x, 3.0 and 3.0.1 (August 2004 update) are not compatible with each other and cannot be converted downward. Therefore, when exchanging documents, it is crucial that you ensure that all parties working on the documents use exactly the same InDesign version. Adobe has announced downward compatibility via the InDesign Interchange Format for the recently-released 4.0 version.
Missing fonts and image files are a second frequent problem. It is necessary not only to keep track of all fonts and EPS files used in the document, but the files used to create the EPS files as well. It is advisable that the DTP provider/department deliver the complete file set together with the final document. Collecting them later may turn out to be a time-consuming and rather costly effort.
The third issue can also be the most frightening: after importing the translation into InDesign documents, parts of the text might be misplaced, there may seem to be disappeared graphics overlapping text paragraphs, the formatting becomes corrupt and TOC entries are missing. Normally, these problems are not caused by the export/import process itself, but rather stem from the original document layout. Everyday experience at milengo has shown us that InDesign documents sent to us for translation may use insufficient or inconsistent layout techniques. An example would be TOC entries that have been added manually to the automatically generated TOC without applying appropriate paragraph styles to the respective headlines in the text body. Although doing so makes the text visible in the printed version or PDF document, the automatic TOC generation would not compile correctly after translation and would therefore need to be manually corrected again. Displaced or invisible text typically occurs if the frames in the original InDesign document do not leave any space for increased text length in other languages.

To avoid any potential predicaments that may stem from the layout, it is highly recommended to have localization in mind as early as possible when creating InDesign document. The next chapter suggests some layout guidelines for "localization-ready" InDesign documents. If documents are to be translated into multiple target languages, it might even be useful to adjust source documents to meet these guidelines prior to exporting the translatable text. Doing so can significantly reduce additional adjustments and alterations of each of the target documents after translation.

Table 1. Filter Tools for InDesign CS – Feature Overview

	TRADOS Story Collector	SDLX	STAR Transit	ECM SysFilter
Version tested	LSP 6.5.5.438	SDLX 2004 Bonus Release	Transit XV Build 518	Build 11
Exported file format	ISC	ITD	TTC	RTF
Exported text type	Tagged text, proprietary	Tagged text, proprietary	Proprietary	Formatted text [1]
Translation compatibility	TRADOS Tag Editor	SDLX, TRADOS Tag Editor	STAR Transit	Any RTF-compatible editor
Tag safety	(+)	(+) [2]	(+)	(+) [1]
Handling of inline image frames	(~) [3]	(+)	(+)	(-) [4]
Handling of hyperlinks/TOC/Index	(-)/(-)/(-)	(~) [5] /(+)/(+)	(+)/(+)/(+)	(~) [5] /(~)/(~)
Handling of index marks	(~)	(~)	(+)	(+)
Manual copy and paste	(-)	(~)	(-)	(+)
Splitting of large translatable files	No	No [6]	No	Yes
Filter works with InDesign 2.x/3.0/3.0.1	Yes/Yes/Yes	No/Yes/Yes	Yes/Yes/Yes	Yes/Yes/Yes

(-) bad, (~) poor, (+) good
[1] Tags used to separate stories only (in “Export to Word“ mode)
[2] Translators cannot remove tags
[3] Import feature places inline frames at the start of story
[4] Inline frames are stripped from InDesign file during import
[5] Hyperlinks are stripped from document during import
[6] Splitting feature is available with SDLX 2005

Layout Guidelines for Preparing InDesign Documents

- Use paragraph styles to format text elements.
- Use character styles only if appropriate.
- Make full use of InDesign automation features like TOC and index generation.
- Use a page layout that will not be disrupted by increasing or decreasing text lengths or frame sizes. It is ideal that some space be left on each page.
- Enlarge text frames (including image labels) to ensure adequate space for potential increase in the text length in the target language.
- Unlink text frames unless there is a good reason to link them. This prevents the target text from "jumping around" in the target documents, which may be in a language that the technician does not understand. ECM SysFilter automatically unlinks text frames, but the assignment of text to frames should be approved.
- If you do not want to unlink frames, consider using frame-break characters to separate text elements that should appear in different frames.
- Always activate wrapping or transparency for foreground frames overlapping text frames.
- Split large InDesign files into multiple smaller files to limit the impact of potential export/import problems. Be sure to make use of the InDesign book function to ensure consistent page numbering and index creation.

Workflow Aspects and Quality Assurance

Localization can be greatly simplified if potential localization pitfalls are considered during the initial document creation. First of all, the graphic design agency or department, the LSP and the client should agree as to which InDesign version will be used. If you use layout specifications, take localization issues into account. All fonts and image files should be stored safely.
Before starting the actual localization of large documents, perform a pilot test by running a small portion through the entire localization process for each target language. This is to make sure that all document and language-specific technical issues are identified and addressed as early as possible.
While the translation will usually be performed in the target-language countries, it is highly recommended that you perform all conversions and layout adjustments at a central location. The party that exported the translatable text should be responsible for importing it back again.
Localized documents should be approved by both the LSP and the client. The DTP staff in charge of QA should be provided with a PDF export of the localized document so that they can record and track their changes by adding comments. Requested changes should be implemented at the same site responsible for all other document conversions and adjustments.

Conclusion

All available tools are capable of exporting and importing the main text threads, but only the Transit filter can fully convert all relevant InDesign features. Unfortunately, Transit only works with a proprietary file type that cannot be employed with any other translation tool. Trados or SDLX users may likely choose their filter based on the relevant features required for the particular document or project. For example, if there are several inline images to process, SDLX would be the preferred filter. If translators have to fiddle with complex tables, or if compatibility is a priority, ECM SysFilter would be the best choice.
As within any aspect of the entire localization process, careful pre- and post-processing plays a crucial role in the localization of InDesign documents successfully. Whatever tool you decide to use, the document layout should ideally be developed with localization in mind, and a well-defined workflow with your localization provider should be defined to significantly reduce any unnecessary manual work on behalf of the DTP engineer. Therefore, when outsourcing InDesign localization projects to a localization agency, make sure they not only prove their localization competence, but that they also have in-depth expertise in InDesign and desktop publishing. We therefore recommend agencies and clients engage in pilot testing phases in the early stages of a relationship to help avoid any pitfalls in the localization process.