

# Translation of Magazines and Brochures Using a Tagged RTF File Workflow

David Calvert  
TransForm Gesellschaft für Sprachen- und Mediendienste mbH  
Cologne, Germany  
[d.calvert@transformcologne.de](mailto:d.calvert@transformcologne.de)

## 1 Introduction

The translation of corporate magazines and brochures poses a particular set of challenges due to the need for specific expertise in the areas of DTP software, typography, reprographics and hardware platforms. Integrating translation memory and terminology databases while making widespread use of freelance translators further complicate the process. This paper looks at these issues and how they have been tackled in one company.

### 1.1 Approaches to the multilingual publication process

The old-fashioned approach involves the customer sending a printed copy or a pdf file of the finished original-language edition, which is translated into a Word document and sent off to the client, who passes it on to the design agency. One or more correction cycles, each of which usually involves semi-legible faxes then follow. This approach, which keeps responsibility for the DTP work completely in the hands of the original designers, is often preferred on organizational or contractual grounds. The disadvantages inherent in this process are duplication of effort, relatively high costs, an inbuilt tendency to introduce linguistic and typographic errors and, from the translator's point of view, the barriers it imposes to the use of translation support tools such as translation memory and automatic terminology management. This process is still used by some of our customers, although most of them have moved on.

An alternative approach is for the translation work to be done directly on copies of the source-language layout files, either on the customer's, the design agency's or the translation company's premises. This approach is the most flexible, but requires tying up expensive DTP hardware and software along with highly trained design and translation staff.

A combination of these two methods, with preliminary translations being done in Word and most of the final translation and copy-editing work being done on DTP equipment by translators and/or editors with help from DTP personnel is one way of solving the problems involved — particularly with publications that are subject to substantial revision up to almost the moment of publication. However, this still entails relatively high costs. We translate regular magazines for customers who require this degree of flexibility and accept the associated financial disadvantages.

None of these translation workflows, however, is as cost-effective and efficient for the majority of magazines and brochures as an approach that decouples the translation and DTP tasks.

### 1.2 Decoupling translation from DTP

There are many reasons for decoupling the translation process from the DTP software used for magazine production.

The expense and complexity of the DTP hardware and software alone mean that few agencies, let alone freelancers, can afford to use such equipment for typing in translations. Few translators either have, or are interested in acquiring, the expertise required to operate the software. In addition, the word-processing features of the most popular packages are much less powerful and sophisticated than those of Word-processing packages such as MS-Word.

Publishers are also making increasing use of publication systems such as K4. This new development raises additional licensing and Internet connectivity issues, whilst imposing further constraints on the translation process.

By extracting the source layout's text content to one or more files that can be processed on a standard PC running MS Word — preferably without losing the embedded formatting information — it is possible to ensure that much of the translation process can run independently of the DTP process. What's more, the translator can also make use of the commonly available translation support tools. However, this approach raises additional cross-platform compatibility issues, which must be resolved in day-to-day use.

The final part of the overall translation process (the correction and copy-editing phase in the layout) is carried out after the translated texts have been re-imported into the corresponding layouts, and the settings of these layouts adapted for the target language. Our experience is that this phase is best accomplished in close cooperation with the original design agency, so that a uniform corporate look can be maintained across a range of different-language editions, whilst using the typographic conventions appropriate to the respective languages.

We use these methods to handle Latin-alphabet publications on both PC and Mac platforms and work with partners to produce non-Latin script versions of such publications.

## **2 DTP Issues**

### **2.1 Designers and the written word**

Designers tend to treat the written word as just another element in a visual language of colour, tone and form. Their concern with text is mostly limited to visual aspects such as the shade of grey that the text appears to have when set with specific size, weight, leading, track etc. This approach often leads to the use of shortcuts such as using paragraph marks to break lines and produce a nice-looking ragged right edge, or tabs and spaces to produce indents — regardless of the sentence structure of the text. Communicating the fact that someone else is going to be working on the layouts at a later date is usually enough to limit the use of such techniques to acceptable levels. In fact, where a regular publication is concerned, it is often possible to talk to the graphics team and communicate a few essential points about design for multilingual publishing. When initial caps are a major design feature, the designer may have to accept that the same letter just will not work in one or more of the target languages. Similarly, translators may have to find alternatives for some letters at the start of paragraphs. Many designers dislike "I," "J" and "T" as drop or initial caps. This can cause problems in English, given that one of the most common words used at the beginning of a sentence is "The", which, of course, begins with a "T". More complicated typography, such as article titles with different colours and point sizes for emphasizing specific words, is often untranslatable and requires close collaboration between designer, editor and translator to find an acceptable substitute. For example, try fitting the word "environment" (the English translation of the German word "Umwelt") into a green box big enough to hold six letters.

### **2.2 Typography is a function of language**

In the same way that it can be difficult to convince a client with a limited knowledge of the target language of the importance of specific target-language structures, convincing a designer of the need to apply different conventions in the design of the target language edition of a publication can also be something of a challenge.

Examples of typographical issues start at the level of basic punctuation. English uses typographer's quotation marks above the text and directly adjacent to it in the pattern 66 99. French uses guillimets with spaces between the opening « and closing » marks and the quoted

text. German uses either the German system of two commas as an opening mark and the English opening mark to close — or inverted French guillimets, pointing inwards without space between the marks and the quoted text.

English	French	German	German
“ . . . ”	« . . . »	„ . . . “	» . . . «

**Figure 1: Quotation marks**

Or take dashes. English and German both use the hyphen to break words. However, German uses the n-dash to perform a function that is carried out by the m-dash in English.

24-point m-dash	24-point n-dash	24-point hyphen
—	–	-

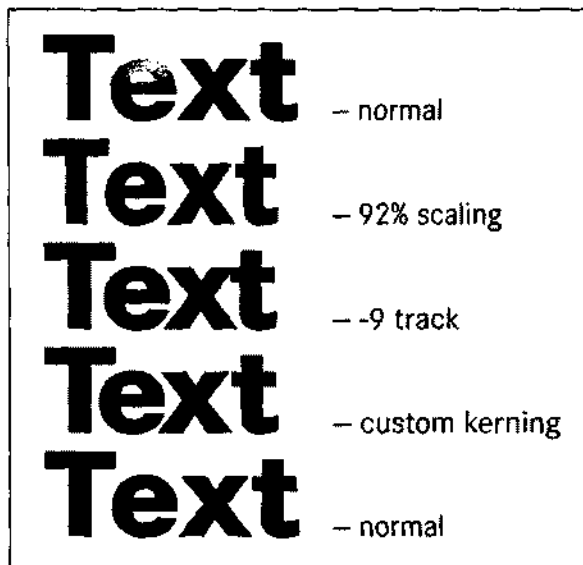
**Figure 2: Dashes**

These points are immediately noticeable in a printed text. But there are also more subtle issues such as hyphenation. The basic algorithms for hyphenation are different in U.S. and UK English, let alone in English and French. The hyphenation settings used in a DTP program typically specify a zone within which words can be hyphenated, how many hyphenated lines may follow one another, a minimum word length and a minimum number of letters at the beginning and end of a word. These "mechanical" parameters interact with the linguistic algorithms and/or dictionaries used to actually split words. As Portuguese or Spanish words, for example, often have more possible hyphenation points than English or German words of similar length, the same "mechanical" hyphenation parameters can produce a different "look" to the text on the page. In fact, the careful choice of these parameters for each language can be employed to minimize the effect of the usual differences in length between them.

Of course, the hyphenation algorithms are not perfect, and some manual hyphenation using discretionary hyphens to force or prevent breaks is almost always necessary.

These typographic techniques can also be used to change the length of a text on the page. Designing a page is always a process of interplay between the written content and the visual aesthetic, with the ideal being to communicate the information in the publication to maximum effect while demanding the minimum conscious effort on the part of the reader. However, it often turns out that the "perfect" text is too long or too short for the space available in the visually optimum layout. There are two frequently used typographical answers to this dilemma. One is to change the horizontal scale of the text by a few percent. The other is to change the spacing between the letters, which is known as the track. These tricks are frequently used on single lines within paragraphs, or on entire paragraphs.

Tracking should not be confused with kerning, which is the spacing between a pair of letters. Designers frequently use custom kerning to make headlines or numbers look better than they would had the automatic kerning been used.



**Figure 3: Scaling, tracking and kerning**

The result is that what looks like a straightforward block of text on the page — at least to the untrained eye — can actually contain a large amount of "hidden" language-specific formatting. This formatting is not only more or less irrelevant to the task of translating; it will negate most of the benefits of a translation memory-based workflow if it is not eliminated.

```

@x: <ip 0, 0, 0, 51, 0, 0, 0, "Dotted2"><z48f"CorpoSExtBol">Text
  <h95z$f"CorpoSReg"><\r> normal
<ip 0, 0, 0, 69, 0, 0, 0, "Dotted2"><h92z48f"CorpoSExtBol">Text<h89, 999>
  <h95z$f"CorpoSReg"><\r> 92% scaling
<t-9h$z48f"CorpoSExtBol">Text<t$> <z$f"CorpoSReg"><\m> -9 track
<z48k-16f"CorpoSExtBol">T<k-10>e<k$>xt <z$f"CorpoSReg"><\m> custom
kerning
<z48f"CorpoSExtBol">Text <h95z$f"CorpoSReg"><\m> normal

```

**Figure 4: Formatting information embedded in the text of Figure 3**

## 2.3 What designers use

### Hardware/software

The sheer variety of possible software, platform and version combinations raises a significant barrier to entry into this sector. A translation company will probably need to have at least one of each of the more frequently used combinations. The vast majority of magazines and brochures published today are produced using either Quark XPress or Adobe InDesign on either Macintosh OS / OS X or Windows platforms. This situation leads to a wide range of cross-platform compatibility issues, which must be resolved in day-to-day use. At least three different versions of Quark XPress and two of InDesign are in widespread use at present on the Macintosh platform. Although the PC platform is not as widely used in the graphics sector as Macintosh systems, it is often very helpful to have the software on a PC as well, if only to help with conversions. Potentially major problems exist when transferring files between versions of the same software and between platforms. In Germany at least, a Mac-based, proprietary ISDN data-transmission system (Leonardo) has, until recently, almost universally been used for file transfer in the graphics and printing sector, which further raises the amount of equipment needed to offer even a basic service in this area.

<b>DTP Application</b>	<b>Version</b>	<b>Operating System</b>	<b>Use by TransForm</b>
InDesign	CS	Macintosh OS X	Frequent
InDesign	CS	Windows 2000/XP	Frequent
InDesign	CS 2	Macintosh OS X	N/A
InDesign	CS 2	Windows 2000/XP	N/A
QuarkXPress	4.x	Macintosh OS 8/9	Frequent
QuarkXPress	5.x	Macintosh OS 8/9	Occasional
QuarkXPress	6.x	Macintosh OS X	Frequent
QuarkXPress	6.x	Windows 2000/XP	Rare

**Figure 5: Common DTP platforms and applications**

### **3 Translation issues**

#### **3.1 What translators use**

Freelance translators use word processing software. Many also use some kind of translation memory system, and perhaps a few electronic dictionaries, both on CD-ROM and online. The fact is that the vast majority of translators use some form of Microsoft Windows application on an Intel or similar platform. This is in stark contrast to the situation in the graphics and DTP sectors where Apple systems are still widely used. Any solution decoupling DTP and translation must therefore enable the error-free interchange of information between the Windows, Mac OS and Unix/Mac OS X worlds.

### **4 Reconciling the two by decoupling**

Decoupling translation from the DTP phases using a tagged rtf file workflow provides a defined cross-platform file format. Although the cross-platform use of rtf files can cause problems, these are not usually significant for our workflow. The rtf files can be used with most translation memory and terminology database systems. This means that we can utilize the substantial translation memories and terminology databases that we have built up for our main customers. These technologies are very useful for maintaining increased terminological consistency across a pool of regular freelance translators and for leveraging previously translated material. We use Trados Translator's Workbench 3 and Multiterm in-house and recommend Wordfast or Trados to our freelancers.

### **5 Basic workflow**

The basic sequence of our workflow is: extract, convert, translate, convert, import and finish.

#### **5.1 A Quark XPress Workflow**

DaimlerChrysler Times is a family of employee magazines published in different editions for the DaimlerChrysler group. The original publications are written in German and, depending on the target readership, translated into English, French, Spanish, Portuguese Turkish and/or Japanese. The format is A3 and the print runs vary from a few hundred to several thousand, depending on the publication and the language.

We receive a copy of the finished German layout in Quark XPress version 6.5 for Macintosh. First of all, we print the job out and prepare pdf files. Then we decide, on the basis of the content, how we are going to divide up the text for translation. Usually, we end up with one file for each page, unless there is a compelling reason for a different split. Because our English translators benefit most from our technology, we often divide up the text on the basis of which stories look familiar to them.

The next step is the extraction of the source layout's text content to one or more files. The formatting information is retained thanks to the use of the "XPress Tags" file format. This format is native to Quark and is supported by the program's filters. It uses a relatively simple system of tags to define the formatting of text within a Quark story.

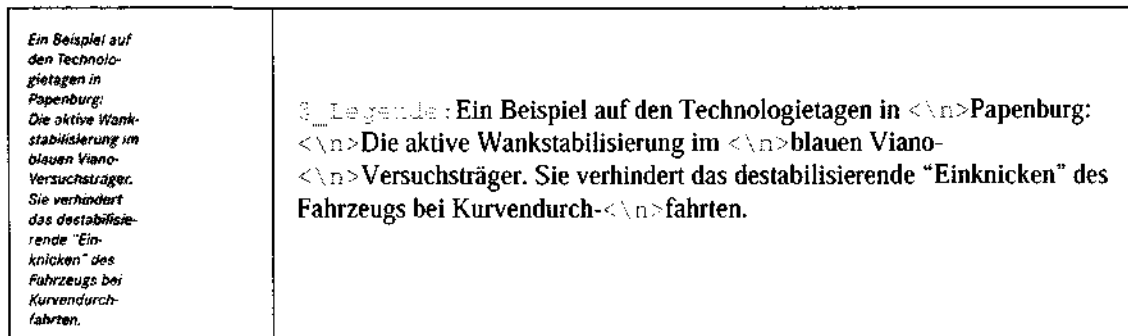


Figure 6: Text and corresponding XPress tags

### 5.1.1 Extraction

To extract the text, we use CopyFlow Gold from North Atlantic Publishing Systems. This is a well-established third-party add-on ("XTension") for Quark XPress, and is available for most versions of Quark on the Macintosh. It enables the user to name stories and select them for export. The XPress Tags format is a text format, and these files must be converted into a format that can be processed on a standard PC running MS Word without losing the embedded formatting information. The Macintosh text files are converted into MS-Word rtf using the program Filter Pack, which has been around in a number of guises for many years. It was made widely available as an extra with Trados Translators Workbench version 3, which is how we came to use it. The specific rtf format is referred to as tagged rtf. It is widely used by TM systems as a simple format capable of representing a wide range of tagged file formats including markup languages and DTP formats. It uses two character styles to mark internal tags, which occur within translation segments, and external tags, which occur outside, and often delimit translation segments. The external tags are normally formatted in grey and the internal tags in red, with the text itself remaining black.

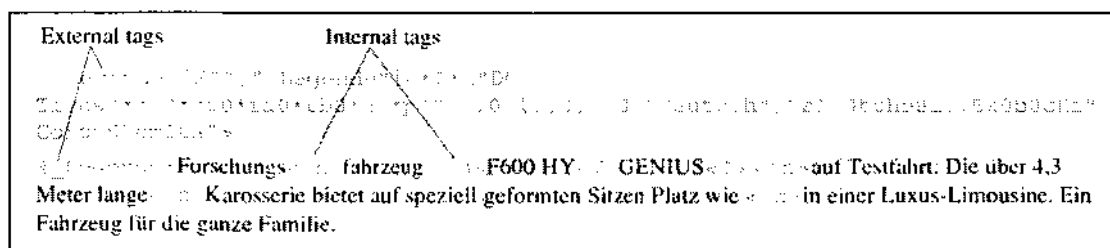


Figure 7: External and internal tags from QuarkXPress

This format can be handled in dedicated editors such as Trados' Tag Editor, or in Word itself. When working in Word, a segmenting translation memory tool such as Trados or Wordfast is helpful, but by no means essential. Such tools take account of the tags and skip the external tags

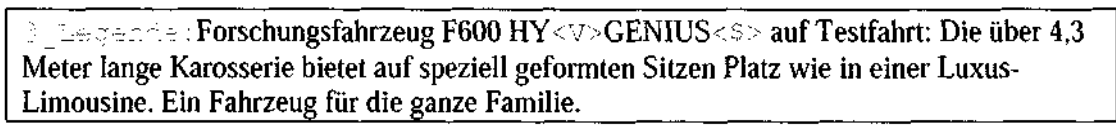
so that the segmented text is presented as text for editing. The internal tags within the segment are treated as placeable elements.

### 5.1.2 Pre-editing

Once the tagged rtf files are available, we find it helpful to edit them prior to translation. Occasionally, it makes sense to carry out a pre-editing step in Quark to fix a badly formed layout, but this is not usually cost-effective. However, such a step is often very useful when it comes to translating long and repetitive order forms or manuals.

Editing the tagged rtf files requires knowledge of the purpose and use of the tags themselves and of their relevance for the translation process. Common problems due to badly formed documents include the use of paragraph marks to break lines, or tabs and spaces to produce indents. These can be quickly and easily eliminated at this stage. Caution is required, however, as any formatting differences — visible here as tags — between fragments of a sentence that has been broken up in this way will cause problems if their tags are deleted.

External tags are, by definition, irrelevant at the translation level. The internal tags, however, are another story. It is obvious that discretionary hyphens are almost entirely irrelevant to the translated text. The same goes for tracking applied to squeeze or expand individual lines. Soft returns are useful in situations such as addresses, but should be removed from paragraphs of body copy, where their job is better done by discretionary hyphens. Searching for and deleting such characters eases the translator's job considerably. But this step cannot be allowed to take up too much time. Our standard approach is to delete all discretionary hyphens, all soft returns in body copy and most track changes affecting individual lines.



3\_Legende: Forschungsfahrzeug F600 HY<v>GENIUS<s> auf Testfahrt: Die über 4,3 Meter lange Karosserie bietet auf speziell geformten Sitzen Platz wie in einer Luxus-Limousine. Ein Fahrzeug für die ganze Familie.

**Figure 8: Paragraph after removing non-essential tags**

There are, however, items of text that use typography for visual effect and do not vary with language. Examples include the superscripted word "GENIUS" in Figure 5, or the kerning of character pairs within a company name or telephone number. It would be counterproductive to eliminate such formatting information only to have to reintroduce it at a later stage. The person carrying out this step must be able to judge which tags to eliminate and which to retain. It is advisable to have a printout of the original layout to hand when carrying out this step.



**Figure 9: The effects of removing too many tags**

### 5.1.3 Translation

At this point, files destined for freelance translators can be presegmented and terminology inserted. We use Trados Team Edition 3 to maintain our translation memories, to presegment and insert terminology in texts sent out to freelancers and, of course, for internal translation. We

encourage our freelancers to use Trados or Wordfast, although it is worth noting that some freelancers prefer to work without the assistance of TM packages, and edit the tagged rtf files directly in Word. We normally send out a pdf of the layout file with the tagged file to eliminate the problem of decontextualization.

Translators working on these files in Word must disable certain features if the files are to be useable for re-import. "Auto-format as you type" is one of the main villains. Revision marking is another, although most translation memory software objects to this feature, and will prompt the user to disable it. The filters can handle straight quotation marks without problems — typographers' quotes, particularly if they are not English, and apostrophes can cause difficulties. Or higher-order characters may not translate correctly to the Macintosh character encoding — the dash can cause problems, depending on the combination of Mac OS version and MS-Word version. In general, the translators have to know enough about the tagged rtf file format to produce "clean" files. Here again, some knowledge of what the tags actually mean is necessary if the maximum gains are to be achieved from the process. We have prepared a short (one page) note for freelancers on the meanings of the most important tags and what to avoid when working with tagged Quark XPress files.

#### **5.1.4 Import**

Once we get the translated files back, they are briefly checked to ensure that there are no obvious problems. These could include comments with terminology inserted during pretranslation that have not subsequently been deleted by the translator, comments inserted by the translator or clearly damaged tags. The files are cleaned up to memory and converted back to the XPress Tags format using Filter Pack. The resulting XPress Tags files are then re-imported into the layouts. Two issues in particular can cause problems at this point. Firstly, it is important to ensure that the type of quotation mark conversion carried out by the file import is set correctly in the Quark XPress preferences. Secondly, it is essential that all of the fonts used by the document are loaded and available to Quark XPress at this point as they will be otherwise be replaced by the default font.

Once the import process is complete, any necessary font substitutions can be carried out and tokens replaced if the target language document needs to be set in an alternative typeface. Later on, I will describe how this is done for Turkish.

If the document is monolingual, it must now be converted to the target language to enable correct hyphenation and spell checking. This is done by setting the format of a block of text to the target language and saving the document as a monolingual document. The program then prompts for the final language. If the target document is to be multilingual, then the individual blocks of text should be formatted with the correct language settings. The most elegant way to accomplish this is to define new stylesheets for each language. Each stylesheet can have a named set of hyphenation parameters associated with it.

#### **5.1.5 Tidying up**

##### **Initial Formatting**

At this point, the format of the target document is probably close to its final form. There are a number of minor issues associated with the various conversion steps, and these require attention before the correction and final editing stages can be started. The main problem associated with this process is the loss of in-line graphics (e.g. as a custom bullet or a keycap in a program manual). There is no alternative here but to copy these from the original. In the case of a bullet, care must be taken to ensure that the associated language format is not inadvertently transferred too. This is because copying the first character of a paragraph in Quark XPress brings the paragraph format with it. Another problem involves formatting that deviates from the base style of the stylesheet. Unfortunately, this is not always correctly retained by the process. The "Lock to Baseline Grid" attribute is particularly prone to this problem.



Once these issues have been dealt with, the scale of the damage to the design becomes clear. Over or under-length text, especially in captions and titles, can be linked into new boxes for correction. We use boxes of the same width as the previous column, with no text wrap and a 100% yellow background to hold overset text. Where the text is reversed we use magenta boxes to make the text clearly visible. Naturally, these temporary modifications have to be deleted before the finished layout is delivered.

### **Line-for-line correction**

At this point, we carry out our line-for-line correction of the translated layout.

We do not normally worry about updating the translation memory with the results of this line-for-line correction. The effort involved is simply not worthwhile given that most of these layout projects do not contain large numbers of 100% matches. And, of course, where the match is only partial, we have to correct the material on a line-for-line basis again anyway.

The exception to this rule occurs in projects where the repetition within the project is obviously high. Examples include closely related order forms or instruction manuals. In such a case, we ensure that the corrections to one part of the translation are incorporated into the memory before proceeding to the next part. In some cases, we create a project translation memory. Where these steps are carried out, it may also be appropriate to pre-edit the source-language layouts to ensure that they are well-formed prior to extracting the text.

### **Final copy-editing**

Once the line-by-line correction stage is finished, the text is edited to ensure that it reads well and fits into the available space in the layout.

Issues such as whether widows (short single lines at the top of a column of text) and orphans (ditto, at the bottom of a column) are permitted, the amount by which text may be too long or too short and the nature and extent of the typographical tricks that can be used to lengthen/shorten the final text must be settled by agreement with the customer and the design agency. This should always be done in advance as they significantly affect the overall cost.

## **5.2 Token-based solution for Turkish**

Where the target language contains characters (sorts) that are not available in standard western European typefaces, special fonts containing these characters must be obtained. This can lead to problems, as different standards exist for Macintosh and Windows implementations of PostScript fonts, True-Type fonts are frequently not acceptable to operators of image-setting equipment, Open Type fonts may not be available for a specific typeface, and Unicode has not yet penetrated far enough into the graphic design universe to eliminate cross-platform character-translation issues. Tokenizing offers a simple workaround, providing that the number of special characters is not too large.

Turkish contains six special characters not found in English. These are shown in Figure 10. Note that the English "I" and "i" belong to different letters in Turkish. This fact is not appreciated by all type foundries, which means that formatting text as "all caps" does not always work correctly.

Ğ	İ	ı	Ş
ğ	ı	ı	ş

**Figure 10: Turkish letters not found in English (upper and lower case)**

The Times New Roman and Arial fonts included with Windows incorporate these characters. Prior to converting the tagged rtf into XPress tags, we run a simple MS-Word macro to replace these characters with tokens. It is important that the tokens use characters that will not interfere with the conversion process. We have found that using "{" and "}" to delimit the tokens works without problems.

{GACC}	{IACC}	i	{SACC}
{CAPGACC}	I	{CAPIACC}	{CAPSACC}

**Figure 11: Tokens used to represent Turkish characters.**

The translated texts, which are imported into the layouts as normal, are initially far too long, due to the tokens. Before this can be remedied, the fonts of the original document must be replaced with the corresponding Turkish fonts. This is easy to do in Quark XPress. Once the correct fonts have been set and activated, the "search and replace" function can be used to replace the tokens with the correct sorts. Thereafter, the process proceeds almost as normal.

There are significant differences between the copy-editing process for languages such as English or French, for which Quark language modules are available and that for languages such as Turkish, where third-party software ("XTensions") must be purchased. Quark XPress Language Modules enable automatic hyphenation of text according to the incorporated rules. In most cases they also incorporate a spell-checking facility. Quark XPress Passport is supplied with a number of such modules in Europe. Supplemental Language Modules are also available from Quark.

When working with languages for which no Language Module is available, hyphenation must be carried out manually or a third-party XTension such as Hyphenator must be purchased for that language. Such add-ons work by writing discretionary hyphens into the entire text according to the "mechanical" hyphenation settings specified in the document preferences. These discretionary hyphens overrule the automatic hyphenation carried out by Quark. It is important to remember that as far as Quark is concerned, the document is still formatted in the original language. If the text is edited, it is necessary to run the hyphenation routine again, otherwise Quark's active language module may automatically hyphenate any new text according to the original — and wrong — language rules. A similar danger arises if the text is copied into a different layout whose preference settings are not identical. On the positive side, the design and setting studios don't need the hyphenation XTension. The same can't be said for the language modules, which must always be present when the document is loaded. This is obviously another area where close cooperation with the design and setting studios is vital.

### 5.3 Non-Latin scripts

To date, we have done relatively little work with non-Latin scripts. We use specialist typesetting services for Russian and cooperate with an agency specializing in translating into and setting Chinese, Japanese and Korean for editions in these languages.

### 5.4 InDesign

During the recent few years, we have received an increasing number of projects that require Adobe InDesign. Here, we have tried with varying degrees of success to use the workflow we have developed for Quark projects. However, the challenges associated with Adobe InDesign appear to be more numerous and more complex.

### 5.4.1 Current InDesign Workflow

Our current, regular projects in InDesign also utilize the publishing system K4. This system stores the layouts in a central database, from which authorised participants can check out layouts or texts to work on. We check out the layouts plus the texts they contain, and save copies of the associated files on our local system. The next stage is to print out the layouts and export screen-resolution pdf files. We extract the text as rtf text without tagging information. These files are tidied up and the number of standard 55-character lines they contain is counted using a Word macro. The texts are then sent for translation in Word (accompanied by PDF files). The translated texts are cut-and-pasted into the InDesign layouts, the layout settings are chosen to suit the target language, and the process of line-for-line correction and final copy-editing is carried out as if we had been working on a Quark XPress document.

### 5.4.2 Issues when working with InDesign

There are a number of reasons why we have not yet adopted a tagged file workflow when it comes to working with InDesign documents. For one thing, until recently, there weren't any suitable tools and filters for extracting and re-importing text into InDesign layouts.

```
Höhe<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000><cT  
racking:-5>punkt war die Feuer- und Lichtschau am Förderturm:  
Pyrotechnische Effekte und ein Feuerkranz am obersten Geländer  
des Gerüsts sowie eine Performance  
<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000><cTrack  
ing:-10>von 75 „Berggeistern“,  
<cHorizontalScale:><cTracking:><cHorizontalScale:1.020000><cTrack  
ing:5>eine Abseilaktion von  
<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000><cTrack  
ing:-5>der 50 Meter hohen  
<cHorizontalScale:><cTracking:><cHorizontalScale:1.020000><cTrack  
ing:8>Plattform, waren alle  
<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000><cTrack  
ing:-10>Teil einer  
Inszenierung<cHorizontalScale:><cTracking:><cHorizontalScale:0.98  
0000><cTracking:-5> der Künstlerin Marianne  
<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000><cTrack  
ing:-3>Manda, die die  
Geburts<cHorizontalScale:><cTracking:><cHorizontalScale:0.980000>  
<cTracking:-5>tagsfeier zu einem außergewöhnlichen Erlebnis  
machten.  
<cHorizontalScale:><cTracking:><CharStyle:Autorenkürzel><cHorizon  
talScale:0.980000><cTracking:-5>Ans  
<cHorizontalScale:><cTracking:><CharStyle:>
```

**Figure 12: InDesign tags**

InDesign documents tend to contain many more tags than a similar-looking Quark document. In addition, the tags are more verbose than those in a Quark document. As a result, the translators working on tagged file in Word are faced with a much more unfriendly picture on their screens. The InDesign import filters that the program uses to re-import tagged text are also much more sensitive to faulty syntax in the tagged files than their counterparts in Quark XPress. This means that removing language-dependent formatting tags without damaging the file beyond use is a much more difficult task, as is translating the text.

One possible alternative might be to use formatted rtf files exported from InDesign. The formatting information in such files is not, however, sufficiently robust to survive the platform and version conversions without corruption. It is thus unsuitable for re-import into the layout. In addition, such files can only be worked on if the appropriate fonts and templates have been

loaded on the computers used for processing. These additional requirements would make it difficult to send work out to freelancers and increase the cost of providing IT support.

On the positive side, cutting and pasting text into InDesign is much faster and easier than doing the same job in older versions of Quark. This is because InDesign can be set to paste text so that it assumes the format of the text at the point of insertion.

InDesign also has good support for scripting. JavaScript, Visual Basic for Applications and AppleScript for Mac OS are all well supported, with documentation and example scripts available from Adobe. Scripts are particularly useful for tasks such as setting the hyphenation or the language throughout an InDesign document. Language is a character attribute in InDesign, rather than a paragraph attribute as in Quark. Globally changing language settings in InDesign — a procedure that would take a long time manually — can easily be done using a script.

We are continuing to look at ways of improving this workflow.

### 5.4.3 Tools for use with InDesign

Leaving aside tools supplied as integral parts of proprietary translation memory suites such as Trados, we have experience of SysFilter for InDesign and CopyFlow Gold for InDesign. Both tools have pros and cons, but they are both capable of exporting text as formatted rtf and as tagged text in the encodings supported by InDesign's filters. Both tools can also import translated texts, subject to the imitations imposed by the Adobe filters. The table below compares some features of the two programs. It does not cover all of the capabilities of either of them.

Tool	Extract formatted rtf	Extract tagged text as text file	Extract tagged text as tagged rtf file	Concatenate stories in a layout	Batch mode	Supplier
<b>SysFilter for InDesign</b>	Yes	No	Yes	Yes	Yes	ECM Engineering
<b>CopyFlow Gold for InDesign</b>	Yes	Yes	No	MS Office macros for rtf files, internal for tagged text	No	North Atlantic Publishing Systems

**Figure 13: Tools for extracting and re-importing text from/to InDesign documents**

SysFilter for InDesign exports Adobe Tagged Text in a tagged rtf format, while CopyFlow Gold for InDesign exports it as pure text. In this case, the text must be converted into a tagged rtf format if it is to be translated using Word on a PC. This can be done using a tagging tool such as Tortoise Tagger (freeware). This tool — in the form of a Word template — uses a list of tags to accomplish the conversion. Tag lists are available for a range of formats, and the author invites users to contribute their own. To date, we have used an in-house Word macro for converting tagged text exported from InDesign. The file specifications for the abbreviated and verbose versions of the Adobe Tagged Text format are available for download from Adobe.

## 6 Conclusions

The use of an rtf tagged file workflow to decouple translation from page layout work in the translation of magazines and brochures brings substantial benefits compared to a more traditional workflow — both in terms of cost savings and improvements in quality. These

benefits are substantial when dealing with documents generated using various versions of Quark XPress and also worthwhile for documents produced using Adobe InDesign. There is scope for further improving the work process associated with InDesign documents — for example, by increasing the use of scripts. Ultimately, however, effective communication between the various parties working on the publications remains the key to a sustainable and cost-effective solution.

The methods I have presented here represent a sensible balance when it comes to satisfying what are often conflicting requirements. The issues involved are complex and range from development and production costs to tight customer deadlines and the limitations of the technology available. More elegant and more effective solutions undoubtedly exist, and we will continue to look for ways to improve the way we deal with such projects.

## **7 References and further information**

*The Chicago Manual of Style*, 14th edition, 1993, The University of Chicago Press

*Hart's Rules for Compositors and Readers at the University of Oxford*, 39th edition, 1983, Oxford University Press

Joy, Lorna, *Translating tagged text — imperfect matches and a good finished job*, in *Translating and the Computer* 24, November 2002, ASLIB London.

Bokor, Gabe, *Translation and Typesetting*, 06/10/2005, *Translation Journal* No. 1, Volume 2, January 1998 (accessible at <http://www.proz.com/translation-articles/articles/334/1/>)

Trados User Guides and documentation, Team Edition 3 <http://www.trados.com/>

Wordfast User Manual, <http://www.wordfast.net>

Tortoise Tagger (freeware) from Aleksandr Okunev at [www accurussian.net/tagger.htm](http://www accurussian.net/tagger.htm).

CopyFlow Gold from North Atlantic Publishing Systems Inc. at <http://www.napsys.com>

SysFilter for InDesign from <http://www.ecm-engineering.de>

Hyphenator from <http://www.techno-design.com>

Quark XPress from <http://www.quark.com>

InDesign from <http://www.adobe.com>

K4 from <http://www.softcare.de>

*All brand names and product names are trademarks or registered trademarks of their respective owners.*