Editing In InDesign And How It Affects Translation

What Happens Behind the Scenes…

INTRODUCTION

While InDesign is an excellent desktop publishing tool for creating content for print or online publishing, in its core, it has a few flaws that affect professional translation and make translation leveraging from one version to another difficult.

If you develop InDesign content that will be translated into foreign languages then read on. This paper will explain what happens in the background and why this matters when it comes to professional translation.

KEYWORDS: *translation memory, segmentation, fuzzy matching, file preparation, translation, translation leveraging*

CONTENTS

WHY INDESIGN?

1. Because it is a great tool.
2. Because Adobe has captured more than three quarters of the desktop publishing market.
3. Because people looking to write and present their content efficiently and professionally are not going to give it up.

These reasons are good enough to justify finding a way to work around the tool's glitches and find a solution for professional translation and efficient translation updates.

SEGMENTING TEXT FOR TRANSLATION

Before we start talking about InDesign, let us spend a few moments on explaining how a typical translation memory tool works and how it segments text for translation[1].

SENTENCE-BASED SEGMENTATION

Simply said, Translation Memory is a database of original and translated segments that are paired. These translated segments can be reused in any future translation where identical or similar segments exist. In most cases, the most practical segmentation is defined at the sentence level. *One sentence = one segment.*

Let's use this example:

> The software you just purchased runs on a number of supported platforms. See your product's system requirements for the list of supported platforms. Equally, you can visit our website for further information on hardware and software system requirements.

Ideally, the above text would be segmented on a sentence level and broken down to three segments:

> **1.** The software you just purchased runs on a number of supported platforms. **|**
> **2.** See your product's system requirements for the list of supported platforms. **|**
> **3.** Equally, you can visit our website for further information on hardware and software system requirements. **|**

---

[1] For a more detailed explanation download our white paper *Is "Fuzzy Matching" Still Fuzzy to You?* available on our website.

Unfortunately, this is not the case because of the way InDesign applies the code in the background. If you are in the editing view of InDesign all looks fine. There are three sentences. However, when presented in the text format you will see that InDesign inserts a variety of tags.

The following is the same example as above but displayed as text.

```
<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/[No character style]"
FontStyle="55 Roman" PointSize="8" KerningMethod="$ID/Optical" FillTint="90">
        <Properties>
                <Leading type="unit">11</Leading>
                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>
        </Properties>
        <Content>The software you just purchased runs on a number of supported platforms.
See </Content>
</CharacterStyleRange>
<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/Hyperlink" FontStyle="55
Roman" PointSize="8" KerningMethod="$ID/Optical" FillTint="90">
        <Properties>
                <Leading type="unit">11</Leading>
                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>
        </Properties>
        <HyperlinkTextSource Self="uaa7" Name="Hyperlink 2" Hidden="false"
AppliedCharacterStyle="CharacterStyle/$ID/Hyperlink">
          <Content>your product's system requirements</Content>
        </HyperlinkTextSource>
 </CharacterStyleRange>
<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/[No character style]"
FontStyle="55 Roman" PointSize="8" KerningMethod="$ID/Optical" FillTint="90">
        <Properties>
                <Leading type="unit">11</Leading>
                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>
        </Properties>
        <Content> for the list of supported platforms. </Content>
        <Br />
        <Br />
        <Content>Equally, you can visit our website for further information on hardware and
software system requirements.</Content>
        <Br />
        <Br />
        <Br />
</CharacterStyleRange>
<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/[No character style]"
FontStyle="55 Roman" PointSize="8.8" KerningMethod="$ID/Optical" FillTint="90">
        <Properties>
                <Leading type="unit">11</Leading>
                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>
        </Properties>
        <Br />
</CharacterStyleRange>
```

A regular translation memory tool will not know how to process the tags and, as a result, the segmentation will look like this:

1. The software you just purchased runs on a number of supported platforms. See **|**
2. your product's system requirements **|**
3. for the list of supported platforms. **|**
4. Equally, you can visit our website for further information on hardware and software system requirements. **|**

FORMATTING

The translation memory stores the text along with its formatting in a text form. This is very useful because not only we leverage the translation but the formatting as well, thus minimizing the post-translation layout work.

This also means, however, that only those segments where both the content and formatting are identical will be a 100% match, i.e. a sentence that can be fully reused.

In the cases where the content is the same but the formatting is different, the translation memory will apply a penalty and you will only have a fuzzy match.

So, ideally, if you don't change the content then you should not change the formatting either.

INCREMENTAL CHANGES IN INDESIGN

But how do you know if you have actually made changes or not?

One of the main problems of InDesign is the way it records useless incremental changes. InDesign seems to record every action a designer takes, even if it does not change the layout.

For example, one of our clients presented us with an InDesign file containing the phrase: "Professionally Managed by Experts:". In the editing view of InDesign it looked perfectly fine. However, in the text-based format it looked like this:

```
<Content>Profes</Content>

</CharacterStyleRange>

<CharacterStyleRange
AppliedCharacterStyle="CharacterStyle/$ID/[No character style]" FillColor="Color/PANTONE
376 C 2" FontStyle="77 Bold Condensed" PointSize="8.8" Tracking="-10" FillTint="100">

<Properties>

<Leading type="unit">12</Leading>

<AppliedFont type="string">Helvetica Neue LT
Std</AppliedFont>

</Properties>

<Content>sionally Managed by Experts:</Content>
```

It seems that while editing the writer must have selected "Profes" and applied some style which turned out to be the same as the rest of the phrase. This change was thus not visible but yet recorded by InDesign.

5

RESULTING TRANSLATION DIFFICULTIES

The more you manipulate/edit an InDesign document the "dirtier" it becomes. The usual translation memory tools don't know how to handle this "dirty" code and will break the text down to segments that are not logical.

This has two main disadvantages and therefore cost implications:

DIFFICULTIES WITH TRANSLATION

When translators translate within the translation memory environment, they are presented with the individual segments – one at a time. If the sentences or even words are incomplete it becomes difficult to comprehend and thus translate. For languages that have different syntax it may become impossible.

Consider an earlier example from this paper. How would you translate these two segments?

Segment #1: *"Profes"*
Segment #2: *"sionally Managed by Experts:"*

LOSS OF TRANSLATION LEVERAGE

When you translate a document, you want to preserve as much of its translation for future reuse as possible. So when you have a translation memory database that contains a number of translated sentences then you can reuse these sentences if they exist in your updated document.

However, if the segments change, you either have a low-level match (in the best case scenario) or no match at all and the sentence needs to be translated from scratch.

Consider the same example. In the original text this sentence was intact:

Intact segment: "Professionally Managed by Experts:"

If you introduce additional formatting in your next version (albeit inadvertently) then the phrase will be broken to two segments:

Segment #1: *"Profes"*
Segment #2: *"sionally Managed by Experts:"*

You will no longer have a match in the translation memory.

6

EZGLOBE'S SOLUTION INCREASES THE LEVERAGE

In the ideal world, InDesign would have a "clean my document" function. Unfortunately, this is not the case.

At EzGlobe, we constantly develop and improve our own set of tools that complement standard commercial computer-aided translation technology. One such tool is our programmable parser that can be adjusted based on various criteria and can better organize the content requiring translation.

In the case of InDesign, we use this tool to clean up the code and convert it to something simpler and more digestible by translation memory. Concretely we merge the content that was originally broken down by InDesign. While we cannot always achieve full matches, at least we create segments that are closer to those stored in the memory which are considered fuzzy or partial matches.

The following example shows how our parser cleans up the InDesign code.

BEFORE	AFTER

`<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/[No character style]" FontStyle="55 Roman" PointSize="8" KerningMethod="$ID/Optical" FillTint="90">`
`        <Properties>`
 	
`                <Leading type="unit">11</Leading>`
`                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>`
`        </Properties>`
`        <Content>`The software you just purchased runs on a number of supported platforms. See `</Content>`
`</CharacterStyleRange>`
`<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/Hyperlink" FontStyle="55 Roman" PointSize="8" KerningMethod="$ID/Optical" FillTint="90">`
`        <Properties>`
`                <Leading type="unit">11</Leading>`
`                <AppliedFont type="string">Helvetica Neue LT Std</AppliedFont>`
`        </Properties>`
`        <HyperlinkTextSource Self="uaa7" Name="Hyperlink 2" Hidden="false" AppliedCharacterStyle="CharacterStyle/$ID/Hyperlink">`
`            <Content>`your product's system requirements`</Content>`
`        </HyperlinkTextSource>`
`  </CharacterStyleRange>`
`<CharacterStyleRange AppliedCharacterStyle="CharacterStyle/$ID/[N`
	`<tu>` The software you just purchased runs on a number of supported platforms. See `<tag>` your product's system requirements `<tag>` for the list of supported platforms. `</tu>`
	
	`<tu>` Equally, you can visit our website for further information on hardware and software system requirements. `</tu>`

```
o character style]" FontStyle="55 Roman"
PointSize="8" KerningMethod="$ID/Optical"
FillTint="90">
		<Properties>
			<Leading
type="unit">11</Leading>
			<AppliedFont
type="string">Helvetica Neue LT
Std</AppliedFont>
		</Properties>
		<Content> for the list of supported
platforms. </Content>
		<Br />
		<Br />
		<Content>Equally, you can visit our
website for further information on hardware
and software system requirements.</Content>
		<Br />
		<Br />
		<Br />
</CharacterStyleRange>
<CharacterStyleRange
AppliedCharacterStyle="CharacterStyle/$ID/[N
o character style]" FontStyle="55 Roman"
PointSize="8.8" KerningMethod="$ID/Optical"
FillTint="90">
		<Properties>
			<Leading
type="unit">11</Leading>
			<AppliedFont
type="string">Helvetica Neue LT
Std</AppliedFont>
		</Properties>
		<Br />
</CharacterStyleRange>
```

CONCLUSION

InDesign is an excellent desktop publishing tool for creating content for print or online publishing because of its WYSIWYG (what you see is what you get) design interface. However, the more you manipulate/edit an InDesign document the "dirtier" it becomes in its core and you will have less leverage from one version to another.

While EzGlobe's approach helps to "clean up" the file and increase the leverage, designers need to be careful when introducing changes in InDesign as not to "pollute" the code.