PROGRAMA DOCTORADO "LENGUA Y LINGÜÍSTICA INGLESAS"

Lost in XLIFF translation

XLIFF effect on translation process

Mª Elena de la Cova Morillo-Velarde
Curso académico 2009- 2010

FACULTAD DE FILOLOGÍA

PROGRAMA DOCTORADO "LENGUA Y LINGÜÍSTICA INGLESAS"

LOST IN XLIFF TRANSLATION

XLIFF EFFECT ON TRANSLATION PROCESS

TRABAJO DE INVESTIGACIÓN DE

Mª ELENA DE LA COVA MORILLO-VELARDE

DIRECTORA: PROF. DRA. GABRIELA FERNÁNDEZ DÍAZ

AÑO ACADÉMICO 2009-2010

TABLE OF CONTENTS

LIST OF FIGURES 3

ACKNOWLEDGEMENTS...6

1. INTRODUCTION..8

2. AIM AND SCOPE OF THE RESEARCH 10

3. THEORETICAL BACKGROUND..12

3.2. GILT 12
3.3. XML..14
3.4. XLIFF 18
3.4.1. Localisable objects...20
3.4.2. Supplementary information...22
3.4.3. Administrative information 24
3.4.4. Custom data 24
3.5. CAT TOOLS...25
3.6. TRANSLATION OF MARKUP LANGUAGE TEXTS 32
3.6.1. WYSIWYG 32
3.6.2. Text Editors...33
..33
3.6.3. Translation Memory Tools..34
3.6.4. Tag Editors..34
3.7. XLIFF EDITORS 35

4. CASE STUDY..40

4.1. METHODOLOGY 40
4.2. XLIFF ANALYSIS..46
4.2.1. Original file...47
4.2.2. XML declaration 49
4.2.3. Headings 50
4.2.4. Images and captions 53
4.2.5. Lists..55
4.2.6. Footnotes...60
4.2.7. Links...63
4.3. QUESTIONNAIRE 65
4.4. PROPOSAL..68

5. CONCLUSION AND FUTURE WORK...74

6. REFERENCE LIST AND BIBLIOGRAPHY 78

BOOKS AND JOURNALS 78

ELECTRONIC RESOURCES 79
APPENDIX A ....¡ERROR! MARCADOR NO DEFINIDO.
APPENDIX B....¡ERROR! MARCADOR NO DEFINIDO.

LIST OF FIGURES

Figure 1 - XML declaration 16
Figure 2 - XML comments 18
Figure 3 - XML Comments 18
Figure 4 - XLIFF structure 19
Figure 5 - File element 20
Figure 6 - Trans-unit element 21
Figure 7 - Alt-trans and phase elements 21
Figure 8 - Skeleton 23
Figure 9 - Inline elements 23
Figure 10 - Administrative information 24
Figure 11 - Custom data 24
Figure 12 - Exact matches 28
Figure 13 - Full matches 28
Figure 14 - Fuzzy matches 29
Figure 15 - Term matches 29
Figure 16 - Trados Workbench 30
Figure 17 - Idiom WorldServer 30
Figure 18 - Gtranslator 31
Figure 19 - WYSIWYG 33
Figure 20 - Text editors 33
Figure 21 - HTML in SDL Trados Studio 34
Figure 22 - Trados TagEditor 35
Figure 23 - Open Language Tools XLIFF Translation Editor 1 36
Figure 24 - TagEditor with XLIFF 36
Figure 25 - XML marker 37
Figure 26 - WordPad 38
Figure 27 - Open Language Tools XLIFF Translation Editor 2 40
Figure 28- Translation Workspace XLIFF editor 41
Figure 29 - SDL Trados Studio 42
Figure 30 - XML editor 43
Figure 31 - GoogleForm questionnaire 46
Figure 32 - Source file in Translation Workspace 47

Figure 33 - Source file in SDL Trados Studio 48
Figure 34 - Source file in jEdit....................48
Figure 35 - XML declaration in Translation Workspace 49
Figure 36 - XML declaration in jEdit....................49
Figure 37 - Headings in source OpenDocument 50
Figure 38 - Headings in Open Language Tools 50
Figure 39 - Headings in Translation Workspace....................51
Figure 40 - Headings in SDL Trados Studio....................51
Figure 41 - Structure codes for SDL Trados Studio....................52
Figure 42 - Headings in jEdit 52
Figure 43 - Image with caption in OpenDocument....................53
Figure 44 - Images/Captions in Open Language Tools....................54
Figure 45 - Images in SDL Trados Studio 54
Figure 46 - Images in Translation Workspace 54
Figure 47 - images in jEdit....................55
Figure 48 - Lists in OpenDocument....................56
Figure 49 - Lists in Open Language Tools....................57
Figure 50 - Lists in SDL Trados Studio 57
Figure 51 - Lists in Translation Workspace 58
Figure 52 - Bullet point list in OpenDocument....................58
Figure 53 - Bullet points in Translation Worskpace 59
Figure 54 - Lists in jEdit 59
Figure 55 - Lists in skeleton....................60
Figure 56 - Footnotes in OpenDocument....................60
Figure 57 - Footnotes in OpenDocument 2....................60
Figure 58 - Footnotes in Open Language Tools....................61
Figure 59 - Footnotes in Translation Workspace 61
Figure 60 - Footnote in SDL Trados Studio....................62
Figure 61 - Footnotes in jEdit 62
Figure 62 - Footnote in Skeleton....................63
Figure 63 - Links in Open Language Tools 63
Figure 64 - Links1 in Translation Workspace....................64
Figure 65 - Links2 in Translation Workspace....................64
Figure 66 - Links in SDL Trados Studio....................64
Figure 67 - Links in jEdit 65
Figure 74 - Header with jEdit....................69
Figure 75 - BPT/EPT inline elements 71
Figure 78.................... ¡Error! Marcador no definido.

ACKNOWLEDGEMENTS

My thanks go to Antoni Oliver, from the Universitat Oberta de Catalunya, for helping me understand the world of XLIFF and always offering his valuable input.

Thanks also to Elisa Alonso for giving me the initial idea for the research project, to María Lema for her feedback and to Irene Moreno for distributing the questionnaire among her network of translators.

Thanks to my director Gabriela Fernández for her precise guidance and her understanding throughout the process.

Finally, thanks to my parents and Henrik for their support.

1. INTRODUCTION

Technical translators often face translation projects in which the source language is not the only challenge. What happens when they have to translate a text in a file format that is unknown to them? What happens when they are sent an XLIFF file for translation?

The topic of this research project is to assess the effect of the XLIFF format in the process of translating software documentation. XLIFF was chosen instead of other markup languages because it was created for localisation purposes and therefore it should be more adequate for translation than other formats. The XML Localisation Interchange File Format (XLIFF) is an XML-based format created to standardise the localisation process. This study aims to assess whether the XLIFF format adds any value to the translation process and whether XLIFF editors make the most of this localisation format.

Translators can potentially receive documents for translation in various formats: HTML, Microsoft Word, XML, TXT, HTML, etc. If the texts are markup languages-based, a new difficulty is added to the translation process: interpreting the tags. Depending on the markup language, tags can be somewhat intuitive or rather unintelligible, which can cause serious problems for the translator. To give an example, the phrase "How to set up your Google account?" in a title should be expressed in Spanish as "Cómo configurar la cuenta de Google" with no question marks, or as a nominal phrase, "Configuración de la cuenta de Google". But if translators do not know if they are translating a heading or a paragraph sentence because they do not understand the tags or because they are not descriptive, then the text will not be translated accurately. Lists of elements will be translated differently depending on whether the list includes elements which are complete sentences or not. Therefore, it is very important that markup languages flag lists of elements somehow; otherwise, they get embedded in the text as if they are paragraph sentences and they are very difficult to identify.

Inaccurate translations and serious quality problems can be caused by lack of information about the layout of the text, lack of context, etc. A possible reason for this can be file formats or tools that do not display the source text layout or context appropriately. On the one hand, providing context and project information is the responsibility of developers and project managers, and should they fail to do so, translators will not be able to see relevant information, even if the tool allows it. On the other hand, the layout information of the source text should be available for translators if the translation tool filters that information correctly.

Translators are often unable to access a considerable amount of information that definitely helps the translation outcome and quality. Therefore, this research project intends to evaluate what kind of information is lost in the translation of XLIFF files and what the reasons for this are.

The research project is composed of this “Introduction” and an “Aim and Scope of the Research” where the main objectives are exposed. Then a “Theoretical Background” section follows which frames the project in its field and explains relevant concepts. The “Case Study” section includes the methodology followed to carry out the Case Study, the Case Study itself, the questionnaire distributed to translators and a proposal for the display of source text information. Finally, the research project offers its “Conclusion and Future Work”.

2. AIM AND SCOPE OF THE RESEARCH

The aim of this research project is:

a) To determine whether the XLIFF format adds value or hinders the translation process.

b) To find out if XLIFF editor tools hide any relevant information for translators.

c) To provide a proposal of how information should be marked by XLIFF editors.

This project assesses the way XLIFF displays source text information. In order to do so, representative corpora have been analysed according to given criteria and in doing this three different translation tools have been used to evaluate how they filter the XLIFF format. Furthermore, a questionnaire has been distributed among translators in order to find out whether they consider XLIFF useful or not and to identify the information they miss in the translation of software documentation and online help.

The initial hypothesis is that translation tools do not provide all the information translators need because they do not take advantage of all XLIFF specifications[1]. However, it needs to be determined whether XLIFF itself provides information in a clear way and whether translation tools use an implementation of XLIFF that enables showing information from the source text.

Translators of software documentation very often translate texts that lack relevant information, and usually the information provided is difficult to understand if the translator does not have a technical background. Translation tools are increasingly more sophisticated and incorporate better functionalities in terms of translation memory and terminology integration; however, they sometimes fail to show clearly basic information such as layout. Consequently, the translation process and quality can be hindered by new technologies when they are not sufficiently sophisticated or when translators do not have the technical knowledge needed to work with new tools and formats.

[1] A specification is an explicit set of requirements to be satisfied by a material, product or service.

On the one hand, developers need to find more effective ways of displaying source text layout in translation tools and of incorporating project information; on the other hand, translators need to acquire markup languages knowledge that will help them understand documentation and online help texts and to interpret tags.

This research project wishes to point out the flaws of the current XLIFF format and editors that affect the translation process and to provide a proposal that could potentially help developers take into account the needs of translators.

3. THEORETICAL BACKGROUND

Globalisation and the technology advances of the last decades have seen the birth of a new translation need: software localisation. The concepts Globalisation, Internationalisation, Localisation and Translation (GILT) need to be described in order to frame the topic of the present research project. This is covered in the "GILT" section.

Software and its related documentation can be contained in various file formats: HTML, DOC, XML, TXT, etc. Certainly nowadays translators can potentially receive documents for translation in all these formats, which implies that the translators need to use different editing tools compatible with all of them and therefore also need to have a high user-level of these tools. Over the years, there have been several efforts made to create one single format which could encompass these various formats and which would meet the needs of the localisation process, but they were complex and lacked functionalities. However, in 2002 XLIFF was created, which meant an important development in the localisation tools industry. XML and XLIFF are described in their corresponding sections in this research project.

As technology has evolved, translation tools have developed as well, having a direct impact on translation quality and translators´ efficiency, in addition to the profile of translators themselves. To better understand how translators interact with the various markup languages and tools it is necessary to summarise the current panorama of translation technology, which is dealt with in the "CAT tools" and the "XLIFF editors" sections.

To sum up, the following sections cover: GILT concepts, XML, XLIFF, CAT tools, translation of markup languages and XLIFF editors.

3.2. GILT

In a globalised world where most people speak or understand English, and contrary to what one might think, people still want to use products in their own languages, which are adapted to their markets. When a new product, a smartphone for instance, is launched in the US, it is expected to be available in other languages and countries almost at the same time as it was released in the US, which requires an incredible localisation and internationalisation effort. Releasing a product simultaneously in different markets is called “simship” (simultaneous shipment).

GILT (Globalisation, Internationalisation, Localisation and Translation) concepts need to be clarified in order to frame the topic of this project.

The Localisation Industry Standards Association (LISA) (LISA, 2010) has provided standard definitions of these four concepts. LISA´s mission is to promote the localisation and internationalisation industry and to provide help for companies that want to globalise their products and services.

Based on LISA´s definitions, Esselink (2000, p. 4) states that:

> Globalization addresses the business issues associated with taking a product global. In the globalization of high-tech products this involves integrating localization throughout a company, after proper internationalization and product design, as well as marketing, sales, and support in the world market.

Thus, to make a product global, a given company needs to set up other services apart from translation, such as a team that can support the product in a certain country and that can look into the legal aspects of it.

“Internationalization is the process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internationalization takes place at the level of program design and document development" (Esselink 2000, p. 2). Internationalisation is crucial for the success of a product in other markets than its home market. To provide an example, a US company launches a website and "just" translates its content into Arabic without adapting the LTR (left to right) writing to RTL (right to left). Arabic readers would probably not have a good impression about the company website and very likely feel that their language is being disrespected. Likewise, a piece of software translated into Japanese would require double-byte character sets[2] instead of single-byte characters so that Japanese characters fit the space provided for the text in the software.

[2] A double-byte character set (DBCS) is a character encoding system that uses one or two bytes. Languages using double-byte character sets are Japanese, Chinese and Korean. (Esselink 2000, p. 468)

Regarding documentation writing, content should also be internationalised, which means that technical writers need to “write for translation" or "write for a global audience". Occasionally the content of software documentation is too US-focused and, as a consequence, examples, explanations and jokes cannot be translated directly. Writers should aim at producing a neutral content that can be easily internationalised.

“Localization involves taking a product and making it linguistically and culturally appropriate to the target locale (country/region and language) where it will be used and sold” (Esselink 2000, p.3). Localisation implies translating the content and also adapting measures, currencies, times, examples, cultural references, etc. In addition to translation, a localisation project includes many other tasks such as project management, software engineering, testing and desktop publishing.

Localisation is mainly related to software and its documentation. This type of content is characterised by the coexistence of natural languages with markup languages or code. Other typical features of localisation are: lack of space to accommodate the target text in a software menu, content embedded in images, placeholders[3], etc.

“Translation is the process of converting written text or spoken words to another language. It requires that the full meaning of the source material be accurately rendered into the target language, with special attention paid to cultural nuance and style” (Esselink 2000, p.4). Therefore, translation is just one of the activities in localisation.

3.3. XML

XML stands for eXtensible Markup Language, which is an initiative proposed by the World Wide Web Consortium (W3C) as an alternative to HTML.

The W3C is an international organisation that develops web standards. This organisation is led by Tim Berners-Lee, who invented the World Wide Web in 1989 and who wrote the first version of HTML (HyperText Markup Language). In October 1994 Tim Berners-Lee founded the W3C, which has as its mission to develop guidelines and protocols that ensure the growth and the right

[3] A placeholder is a symbol that will later be replaced by a string.

functioning of the Web. One of the aspects that the W3C looks into is which approach to follow in order to solve the problems of the technology of the Web. XML is recommended by the W3C as a format for sharing structured information (between programs, between people and between computers and people, both locally and across networks). XML is also recommended as an alternative for HTML and SGML (W3C, 2010).

XML, HTML and SGML are markup languages. Markup languages are languages that "mark up" plain text so it is formatted and displayed in a web browser in interesting and useful ways. Web pages are displayed by browsers when they load a copy of the web documents (plain text files containing markup) (Lee Ford, 2009: 6). Markup languages are characterised by their tags. A tag is a markup construct that begins with "<" and ends with ">". There are three kind of tags: start-tags, for example <section>, end-tags, for example </section>, and empty-element tags, for example <line-break/>.

SGML (Standard Generalised Markup Language) is an international standard of information exchange that uses descriptive markup within a document and that defines three layers (structure, content and style) (Esselink, 2000: 205). SGML is one of the oldest markup languages; it was born in 1986 and is very complex, which is why it is currently hardly used.

HTML, or HyperText Markup Language, is a markup language that was created based on SGML. Its purpose is to describe the structure of web-based documents. HTML is one of the most common formats of displaying pages on the Web, although today XML is preferred by many because it is more flexible than HTML. HTML tags are predefined by a standard DTD (Document Type Definition), which specifies for example that the tag <b> will make the style of the text to bold.

XML stands for eXtensible Markup Language. It was created in 1996 and recommended by the W3C in 1998. XML can be defined as a "metalanguage"[4] because it allows the definition of unlimited languages.

[4] A metalanguage defines the rules and symbols for other computer languages. For example, XML is the metalanguage used for other languages, such as XHTML (Musciano & Kennedy 2007)

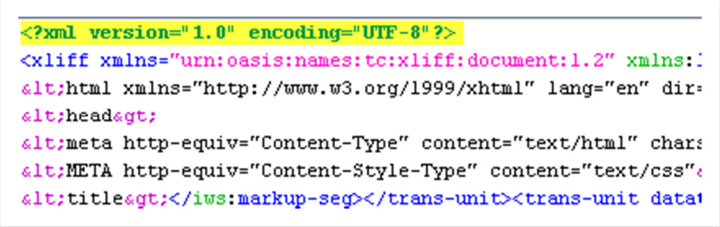
The main purpose of XML is to carry data, not to display data, as HTML does. This means that XML contains information about the text itself; it carries semantic information. For instance, in the following example tags offer certain information about the nature of the text:

```
<Author>

        <Name>Bert</Name>

        <Surname>Esselink</Surname>

        <Nationality>Dutch</Nationality>

</Author>
```

One of the most useful features of XML is that the developers create their own tags, which should be self-descriptive. Contrary to HTML, XML tags are defined by the developer and they depend on the nature and needs of the text.

In terms of the syntax of XML there are several rules that must be followed in order to have a well-formed document:

a) XML documents need to start with an XML declaration, which is the first line of the document. For instance:



```
<?xml version="1.0" encoding="UTF-8"?>
<xliff xmlns="urn:oasis:names:tc:xliff:document:1.2" xmlns:
&lt;html xmlns="http://www.w3.org/1999/xhtml" lang="en" dir=
&lt;head&gt;
&lt;meta http-equiv="Content-Type" content="text/html" chars
&lt;META http-equiv="Content-Style-Type" content="text/css"
&lt;title&gt;</iws:markup-seg></trans-unit><trans-unit data
```

Figure 1 - XML declaration

As the example shows, the XML declaration is composed by the XML version and the encoding used in the document. The question marks need to be written as shown above.

b) XML tags need to open and close, that is, each element needs to contain two tags, one opening the document and another closing it. For instance: <Name>Bert</Name>. If there is not a closing tag, the document will not work. By contrast, HTML documents can have opening tags but not closing tags, and still work properly, as in the example below:

<i> *This text is displayed in italics*

As can be seen, there is no closing tag but the italic type equally works.

The only exception to this XML rule is the XML declaration, which does not contain opening or closing tags.

c) XML tags need to be properly nested. When a tag opens within another tag, it needs to close before the first tag closes. An incorrect example can look like this:

<Book>Song of Salomon<Author>Toni Morrison</Book></Author>

The correct way of nesting these tags is the following:

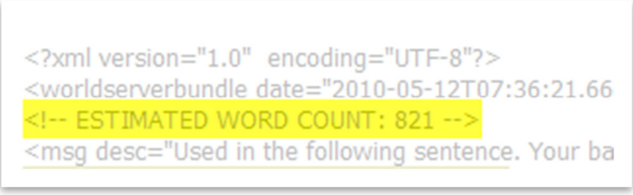
<Book>Song of Salomon<Author>Toni Morrison</Author></Book>

d) XML tags are case sensitive, which means that if the case is different in the opening and closing tag, the document will not work properly. For example, the following text would not work: <Name>Bert</name>.

e) XML tags can contain attributes, which are values of the elements. The attributes from the elements must be quoted, either with single or double quotes. For example: <Book year=”1997”>.

f) Although XML is quite flexible, some elements are considered illegal, such as >, < and “. To represent these elements, XML uses special entity references. The equivalent entity references for these illegal elements (>, < and “) are: &lt;, &gt; y &quot;.

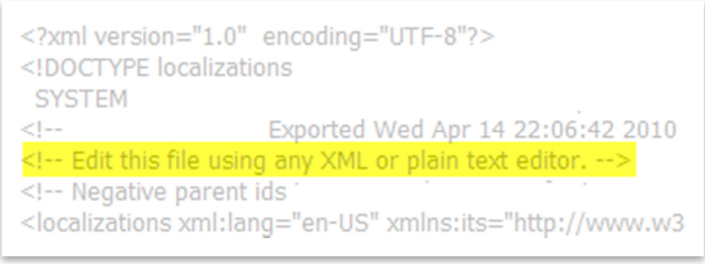
g) Comments also have a fixed way of representation, as can be seen in the word count comment (highlighted in yellow) from figure 2:



```
<?xml version="1.0"  encoding="UTF-8"?>
<worldserverbundle date="2010-05-12T07:36:21.66
<!-- ESTIMATED WORD COUNT: 821 -->
<msg desc="Used in the following sentence. Your ba
```

Figure 2 - XML comments

This comment specifies the word count of the source text, which is a very relevant piece of information for the translator. The developer can include other kind of messages such as the one below, which requires the translator to use an XML editor or a plain text editor to edit the text:



```
<?xml version="1.0"  encoding="UTF-8"?>
<!DOCTYPE localizations
 SYSTEM
<!--                    Exported Wed Apr 14 22:06:42 2010
<!-- Edit this file using any XML or plain text editor. -->
<!-- Negative parent ids
<localizations xml:lang="en-US" xmlns:its="http://www.w3
```

Figure 3 - XML Comments

It is very important for multilingual texts to support non-English letters (á, ó, æ, ø, å…). They are legal in XML as long as the vendor software supports them or the encoding accepts them.

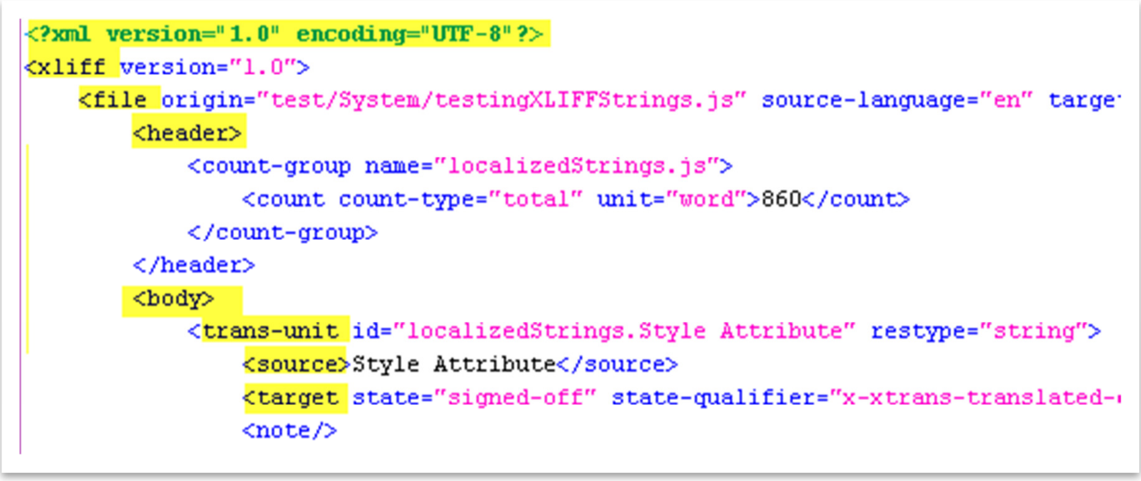
3.4. XLIFF

XLIFF is an XML-based format commonly used today in the localisation industry. XLIFF (XML Localisation Interchange File Format) stores together text and data and carries them from one step to another in the localisation process. It was standardised by the OASIS organisation in 2002. "OASIS (Organization for the Advancement of Structured Information Standards) is a not-to-profit consortium that drives the development, convergence and adoption of open standards for the global information society." (OASIS, 2010)

The XLIFF TC (Technical Committee) charter states:

> "The purpose of the OASIS XLIFF TC is to define, through XML vocabularies, an extensible specification for the interchange of localization information. The specification will provide the ability to mark up and capture localizable data and interoperate with different processes or phases without loss of information. The vocabularies will be tool-neutral, support the localization-related aspects of internationalization and the entire localization process. The vocabularies will support common software and content data formats. The specification will provide an extensibility mechanism to allow the development of tools compatible with an implementer's own proprietary data formats and workflow requirements." (OASIS, 2010)

XLIFF is an XML application and as such, it begins with an XML declaration. After the XML declaration, the XLIFF document itself comes enclosed within the <xliff> element. An XLIFF document is composed of one or more sections, each enclosed within a <file> element. The <file> element consists of a <header> element, which contains metadata about the <file>, and a <body> element, which contains the extracted translatable data from the <file>. The translatable data within <trans-unit> elements is organised into <source> and <target> paired elements. These <trans-unit> elements can be grouped recursively in <group> elements. In addition, XLIFF provides the ability to maintain information about the processing of the file via the <phase> element. These elements are explained in more detail in the following sections. Figure 4 shows an XLIFF file where the basic structure can be seen.



```xml
<?xml version="1.0" encoding="UTF-8"?>
<xliff version="1.0">
    <file origin="test/System/testingXLIFFStrings.js" source-language="en" targe
        <header>
            <count-group name="localizedStrings.js">
                <count count-type="total" unit="word">860</count>
            </count-group>
        </header>
        <body>
            <trans-unit id="localizedStrings.Style Attribute" restype="string">
                <source>Style Attribute</source>
                <target state="signed-off" state-qualifier="x-xtrans-translated-
                <note/>
```

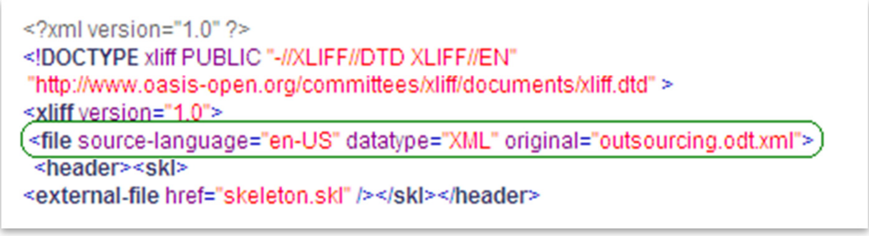
Figure 4 - XLIFF structure

An XLIFF document can capture anything needed for a localisation project:

1. localisable objects (e.g. text strings) in source and target languages
2. supplementary information (e.g. glossaries or material to recreate the original format).
3. administrative information (e.g. workflow data)
4. custom data (e.g. initialisation information for tools).

3.4.1. Localisable objects

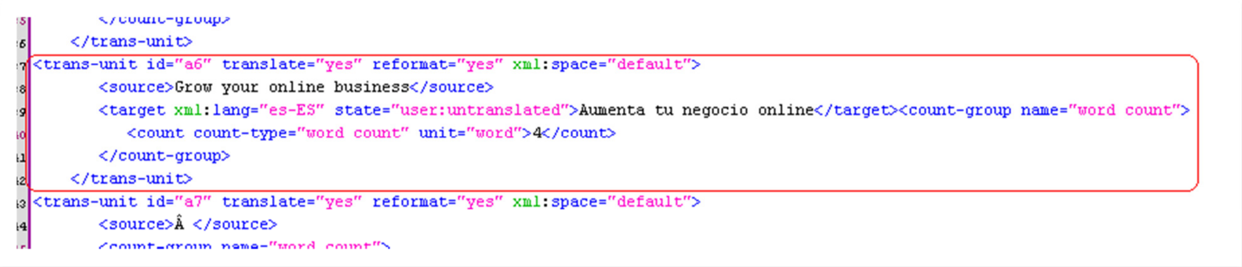
XLIFF can contain the source and the target text. An XLIFF document is composed of one or more <file> elements. Each <file> element corresponds to an original file or source (i.e. database table, file). A <file> contains the source of the localisable data and, once translated, the corresponding localised data for one locale.



```
<?xml version="1.0" ?>
<!DOCTYPE xliff PUBLIC "-//XLIFF//DTD XLIFF//EN"
 "http://www.oasis-open.org/committees/xliff/documents/xliff.dtd" >
<xliff version="1.0">
<file source-language="en-US" datatype="XML" original="outsourcing.odt.xml">
 <header><skl>
<external-file href="skeleton.skl" /></skl></header>
```

Figure 5 - File element

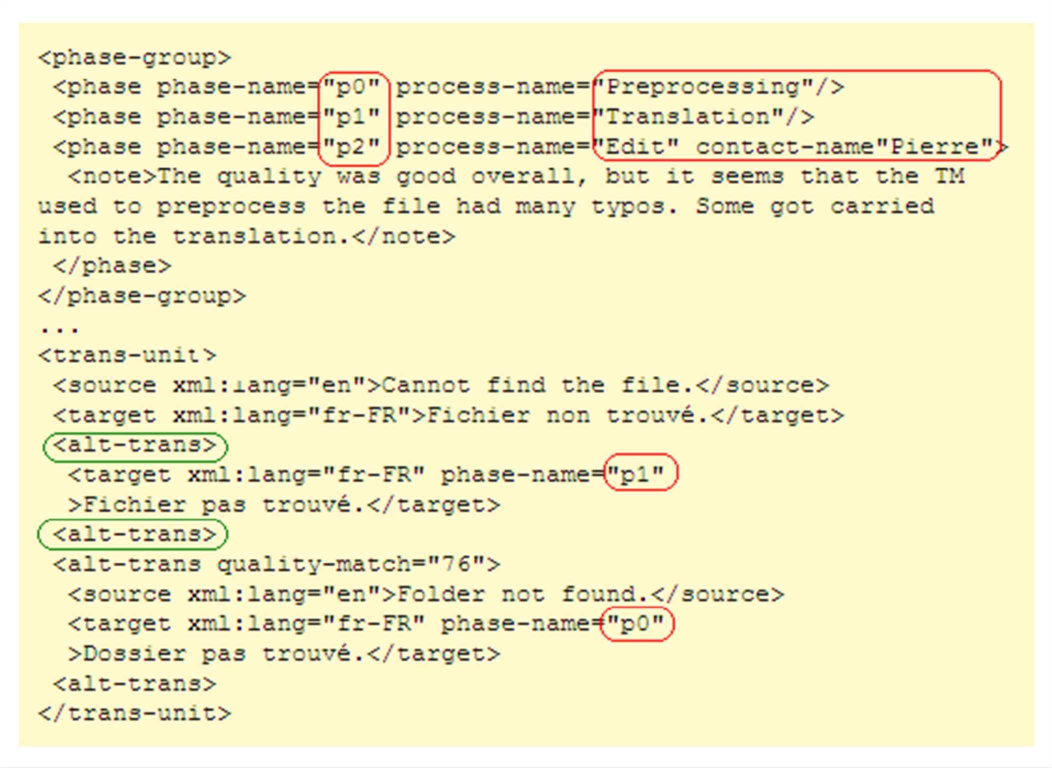
Localisable data are stored in <trans-unit> elements. The <trans-unit> element holds a <source> element to store the source text and a <target> element to store the translated text. For example:



```
        </count-group>
    </trans-unit>
<trans-unit id="a6" translate="yes" reformat="yes" xml:space="default">
        <source>Grow your online business</source>
        <target xml:lang="es-ES" state="user:untranslated">Aumenta tu negocio online</target><count-group name="word count">
           <count count-type="word count" unit="word">4</count>
        </count-group>
    </trans-unit>
<trans-unit id="a7" translate="yes" reformat="yes" xml:space="default">
        <source>Â </source>
        <count-group name="word count">
```

Figure 6 - Trans-unit element

Any translations of the <source> text that are not the latest (i.e. before proof, before edit, etc.), as well as possible proposed translations provided at some point during the localisation process, can be stored in an unlimited number of <alt-trans> elements (see the green rectangles on figure 7). The different versions of the translations can be linked to a specific phase of the process by the phase-name attribute (see red rectangles on figure 7). The details about the different phases can be stored in the <phase> elements, grouped in the <phase-group> element.



```
<phase-group>
 <phase phase-name="p0" process-name="Preprocessing"/>
 <phase phase-name="p1" process-name="Translation"/>
 <phase phase-name="p2" process-name="Edit" contact-name"Pierre">
  <note>The quality was good overall, but it seems that the TM
used to preprocess the file had many typos. Some got carried
into the translation.</note>
 </phase>
</phase-group>
...
<trans-unit>
 <source xml:lang="en">Cannot find the file.</source>
 <target xml:lang="fr-FR">Fichier non trouvé.</target>
 <alt-trans>
  <target xml:lang="fr-FR" phase-name="p1"
  >Fichier pas trouvé.</target>
 <alt-trans>
 <alt-trans quality-match="76">
  <source xml:lang="en">Folder not found.</source>
  <target xml:lang="fr-FR" phase-name="p0"
  >Dossier pas trouvé.</target>
 <alt-trans>
</trans-unit>
```

Figure 7 - Alt-trans and phase elements

3.4.2. Supplementary information

XLIFF allows storing supplementary information, for example, references to glossaries or translation memories. This extra information can be referenced (i.e. reside outside of the document) or embedded within the document:

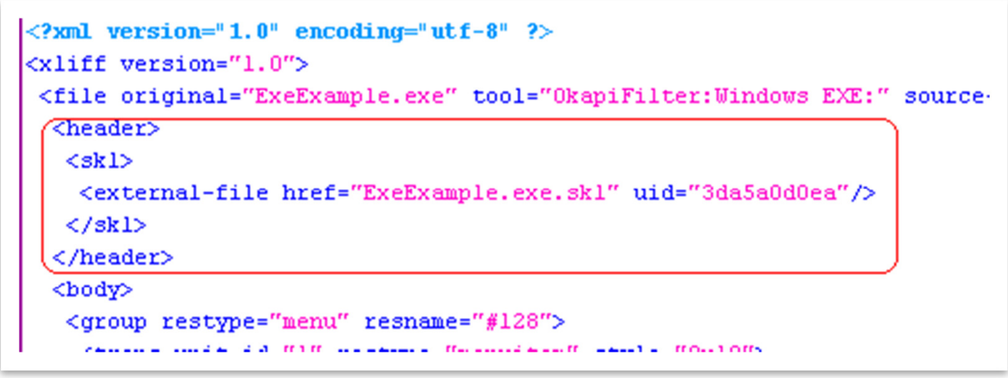
a) Reference to a document outside the file:

```
...<header>
<reference>
 <external-file
  href="TranslationStyleGuidelines.doc"
 />
</reference>...
```

b) Embedded within the document:

```
...<header>
<glossary>
 <internal-file form="text"><![CDATA[
"English term 1","German term 1"
"English term 2","German term 2"
...]]></internal-file>
</glossary>...
```

Non-localisable elements of the file (mainly layout) are contained in the skeleton file, which can be referenced from the XLIFF file or embedded in it. See an example of the reference file:

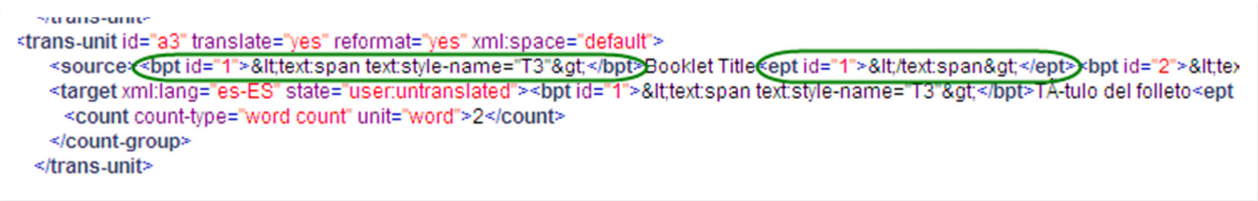


```
<?xml version="1.0" encoding="utf-8" ?>
<xliff version="1.0">
 <file original="ExeExample.exe" tool="OkapiFilter:Windows EXE:" source
  <header>
   <skl>
    <external-file href="ExeExample.exe.skl" uid="3da5a0d0ea"/>
   </skl>
  </header>
  <body>
   <group restype="menu" resname="#128">
```

Figure 8 - Skeleton

The skeleton file (example.skl) and the XLIFF file (example.xlf) make up the zipped or compressed XLIFF export file (example.xlz). To give an example, the file “adsense.odt.xlz” is a zip file made up of “adsense.odt.xml.skl”, which is the skeleton, and “adsense.odt.xml.xlf”, which is the XLIFF file. Some XLIFF editors only work with the XLIFF file (.xlf) and others need the zipped file containing the skeleton (.xlz).

Text layout information is stored in special elements called inline elements, which are used to represent codes that reside within the source or target text, for example the formatting codes to mark a section of a sentence in bold. The inline elements <bpt> and <ept> appear very often in the corpora of this research project. They delimit the beginning of a paired sequence of native codes. Notice that the style of the text is included within the tags.



```
  </trans-unit>
<trans-unit id="a3" translate="yes" reformat="yes" xml:space="default">
    <source><bpt id="1">&lt;text:span text:style-name="T3"&gt;</bpt>Booklet Title<ept id="1">&lt;/text:span&gt;</ept><bpt id="2">&lt;te
    <target xml:lang="es-ES" state="user:untranslated"><bpt id="1">&lt;text:span text:style-name="T3"&gt;</bpt>TÃ-tulo del folleto<ept
      <count count-type="word count" unit="word">2</count>
    </count-group>
  </trans-unit>
```

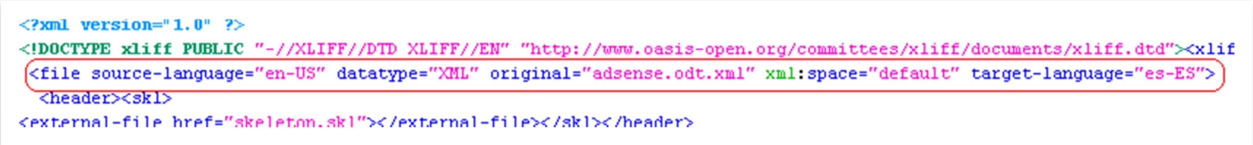
Figure 9 - Inline elements

The complete list of inline elements is: <g>, <x/>, <bx/>, <ex/>, <bpt>, <ept>, <mrk>, <ph>, and <it>. They either delimit or replace code in the original document.

3.4.3. Administrative information

XLIFF provides mechanisms for capturing administrative information such as source material, workflow data, pre-translation entries and keeping track of changes.

Figure 10 shows an example of the source material data:



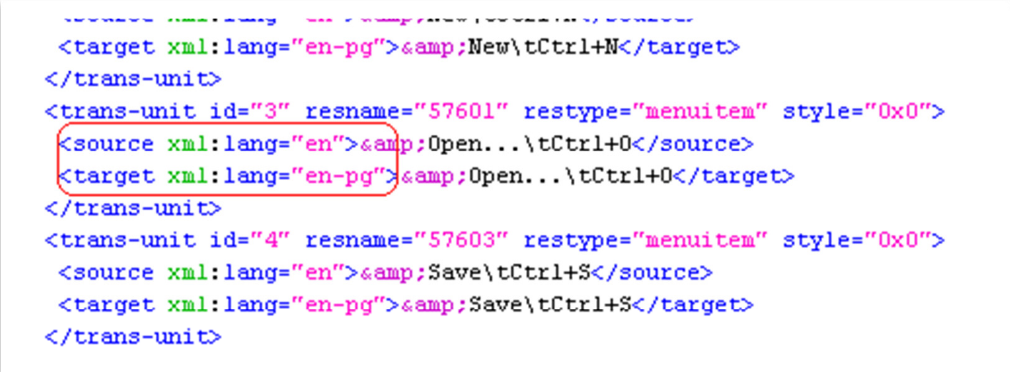
```
<?xml version="1.0" ?>
<!DOCTYPE xliff PUBLIC "-//XLIFF//DTD XLIFF//EN" "http://www.oasis-open.org/committees/xliff/documents/xliff.dtd"><xlif
 <file source-language="en-US" datatype="XML" original="adsense.odt.xml" xml:space="default" target-language="es-ES">
  <header><skl>
<external-file href="skeleton.skl"></external-file></skl></header>
```

Figure 10 - Administrative information

In terms of workflow, track changes and pre-translation, an example was already provided of how they can be marked with XLIFF using <alt-trans> and <phase-group> (see figure 7).

3.4.4. Custom data

Users can customise XLIFF by adding their own elements, attributes, and attribute values. See below an instance of extending element values:



```
  <source xml:lang="en">&amp;New\tCtrl+N</source>
  <target xml:lang="en-pg">&amp;New\tCtrl+N</target>
 </trans-unit>
 <trans-unit id="3" resname="57601" restype="menuitem" style="0x0">
  <source xml:lang="en">&amp;Open...\tCtrl+O</source>
  <target xml:lang="en-pg">&amp;Open...\tCtrl+O</target>
 </trans-unit>
 <trans-unit id="4" resname="57603" restype="menuitem" style="0x0">
  <source xml:lang="en">&amp;Save\tCtrl+S</source>
  <target xml:lang="en-pg">&amp;Save\tCtrl+S</target>
 </trans-unit>
```

Figure 11 - Custom data

As it was mentioned before, XLIFF was born out of the need of having a single standard for the different localisation stages and a format that is compatible with different tools. XLIFF brings considerable benefits to the localisation process for localisation stakeholders such as localisation providers, customers and tool vendors. Some of the main benefits are the following:

- Less dependency on vendors which are able to work with special formats.
- Tighter control on what goes to localisation (pre-filtering of what to translate or not).
- Controlled information flow (author/developer notes, item properties, etc.).
- All advantages of XML-based processing.
- Single format for adjunct processing (e.g. quality control in terms of spell checking).
- Less dependency on specific localisation tools.

3.5. CAT TOOLS

Translation has changed dramatically in the last decades with the implementation of new technologies. Some people still have a romantic idea of the translator as someone spending hours on polishing translations until they are beyond perfect in the target language. Unfortunately, there is no time for that now. The reality is that a professional who wants to live out of translating must translate around 2000 words per day, which is virtually impossible just with a computer and the best human memory.

In addition, the communication between the translator and the client has also been shaped by technology. Email is the main way in which translators and clients make business, receive and send translations. Furthermore, translators need to have a solid presence in the Web to reach clients, and therefore they create profiles in professional networks, become member of translations online forums, associations, etc.

Translation technology has developed quite quickly in the last decades. In 1966, the ALPAC report[5] recommended the development of computer-based aids for translators. Later on, translators were able to get online access to multilingual terms banks, such as Eurodicautom. In 1970, the first proposal for what is now called translation memory was born, and in 1980 Alan Melby proposed the integration of various tools in the translator workstation (Somers 2003, p. 13-14). Nowadays, professional translators use and are required to use what we know as CAT tools.

Computer-Aided Translation (CAT) technology "can be understood to include any computerized tool that translators use to help them do their job" (Bowker 2002, p. 6). CAT tools should not be identified with Machine Translation (MT). Machine Translation applications translate texts, while CAT tools do not; they help the translator translate the text.

Before getting into the details of CAT tools, it is relevant to clarify four concepts related to CAT and MT: Fully Automatic High Quality Machine Translation (FAHQMT), Human-Aided Machine Translation (HAMT), Machine-Aided Human Translation (MAHT) and Human Translation (HT). FAHQMT corresponds to the full use of Machine Translation, HAMT refers to machine activity with some human interaction, MAHT implies human activity with the help of a number of tools (CAT tools) and HT concept denotes human translation without major computer help. Nowadays, HAMT and MAHT are the two types of translations that are most used. The line that separates the four concepts can be very thin and complex, and object of a different research project.

There are different kinds of CAT tools: spellcheckers, terminology managers, WWW, terminology databases, project management tools and translation memories, among others. By far, the most important CAT tool for a translator is the translation memory (TM).

According to Bert Esselink (2000, p. 362):

> Translation memory is a technology that enables the user to store translated phrases or sentences in a special database for local re-use or shared use over a network. Translation memory systems work by

[5] The Automatic Language Processing Language Advisory (ALPAC) wrote a report in November 1964 on progress in Machine Translation research. The infamous ALPAC report concluded that MT was slower, less accurate and twice as expensive as human translation. Although it had some devastating effects for MT, it had some recommendations for machine aids for translators. (Baker 2001, p. 184)

> matching terms and sentences in the database with those in the source text. If a match is found, the system proposes the ready-made translation in the target language.

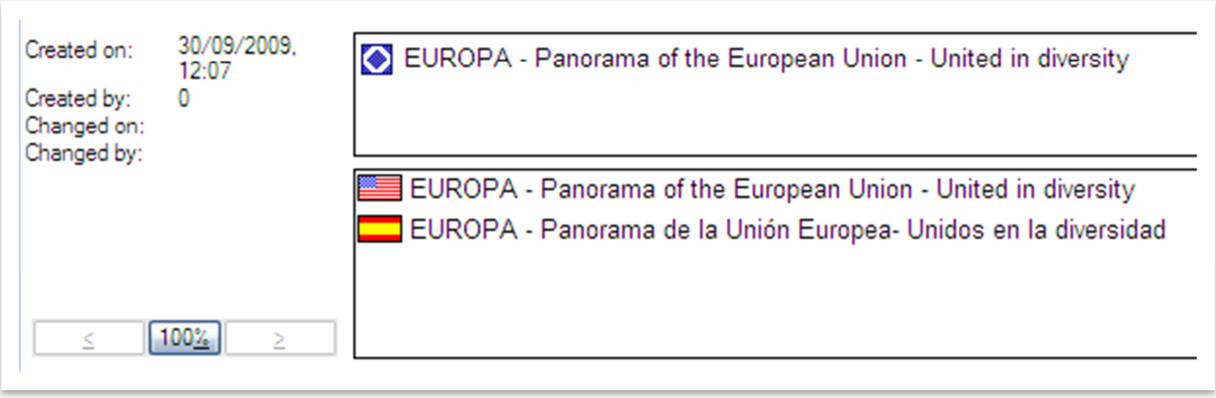
Translation memories are an essential tool for translators nowadays. In the past, translators had to rely on their human memory to translate phrases and terminology that they had already translated in the past. Currently, translators do not have to waste time looking through their stored translations, because translation memories do that for them. Translation memories are important for a variety of reasons. First and most importantly, translation memories save time to the translator. It is believed that a translation memory tool increases productivity by 30%, although that depends on the translation memory itself. Therefore, the more words translators translate the more earnings they will have; especially taking into account that translation is not particularly well paid. A second reason is translation consistency. It is common that a translator works repeatedly for the same client, or that they are specialised in a certain field. Translation memory tools can help them ensure that the same terminology and phrases are used. In particular, this is essential when translators face a large text. As Kay and Röscheisen (1993, p.8) affirm: "One of the most important sources of information to which a translator can have access is a large body of previous translations." Thirdly, translation memories and technology are crucial because translators need to be better skilled in order to be more competitive and more attractive to clients. For instance, in the software translation industry, translators are required to know how to use several tools.

Translation memory technology works by comparing a new source text against a database of texts that have already been translated. When a translator has a new segment to translate, the TM system consults the database to see if this new segment matches a previously translated segment (Bowker 2002, p. 94). The matching of the segments can vary, from 0% to 100% matching. Then, the translator can decide whether to incorporate that segment to the translation or not. In most cases and with a matching over 60-70% the translator will incorporate the segment to the translation and edit it.

Segments are usually sentences, but there are other elements that are also considered segments, such as headings or list elements. Segmentation rules can be defined by the translator.

There are different kinds of TM matches: ICE (In Context Exact), exact, full, fuzzy and term matches.

An ICE match means that the new segment is identical to another in the TM and appears in the same context. It is almost the same as an Exact match, which are 100% identical to the new segment in spelling, punctuation and formatting but do not appear in the same context. Basically, an ICE match is an exact match within the same context.



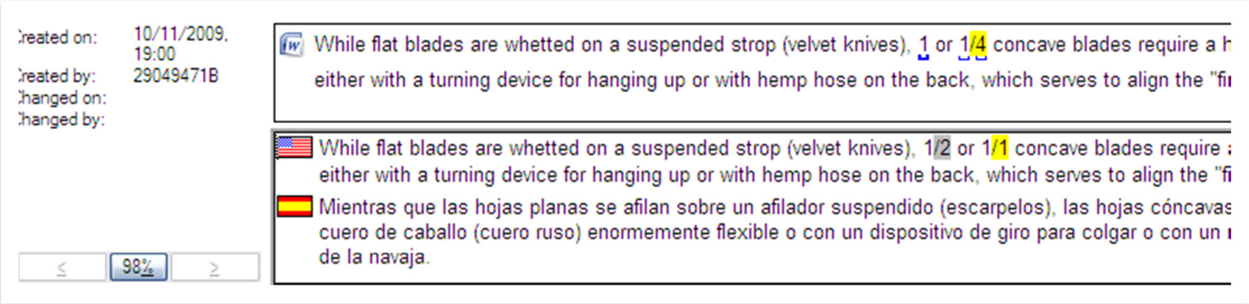
Created on: 30/09/2009, 12:07
Created by: 0
Changed on:
Changed by:

EUROPA - Panorama of the European Union - United in diversity

EUROPA - Panorama of the European Union - United in diversity
EUROPA - Panorama de la Unión Europea- Unidos en la diversidad

Figure 12 - Exact matches

Full matches are quite similar to exact matches; however, there can be differences between the TM and the new segments in what is known as "placeables", which are dates, currencies, measurements or proper names.



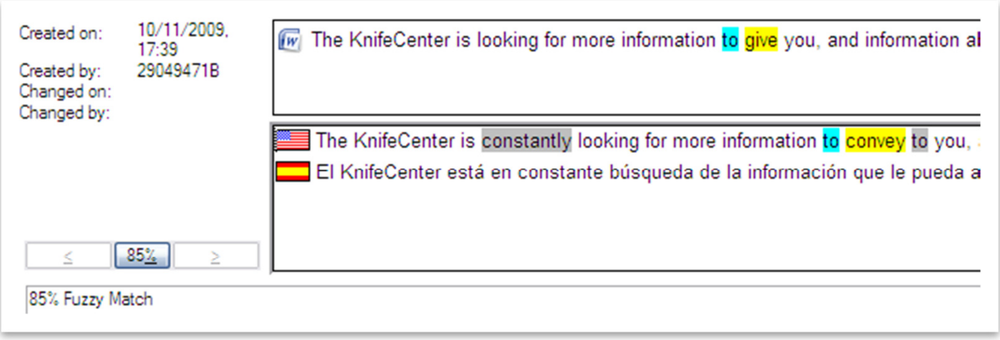
:reated on: 10/11/2009, 19:00
:reated by: 29049471B
:hanged on:
:hanged by:

While flat blades are whetted on a suspended strop (velvet knives), 1 or 1/4 concave blades require a h
either with a turning device for hanging up or with hemp hose on the back, which serves to align the "fi

While flat blades are whetted on a suspended strop (velvet knives), 1/2 or 1/1 concave blades require a
either with a turning device for hanging up or with hemp hose on the back, which serves to align the "fi
Mientras que las hojas planas se afilan sobre un afilador suspendido (escarpelos), las hojas cóncavas
cuero de caballo (cuero ruso) enormemente flexible o con un dispositivo de giro para colgar o con un
de la navaja.

Figure 13 - Full matches

Fuzzy matches retrieve segments that are similar but not identical. Translators can choose where to set the sensitivity threshold, which means that they can decide whether they will specify fuzzy matches to be 90% similar to the new segment or 10% similar to it. Usually, translators set the sensitivity threshold to a 70-75% so that they do not miss potential reusable segments and so that they do not get irrelevant segments.



Created on: 10/11/2009, 17:39
Created by: 29049471B
Changed on:
Changed by:

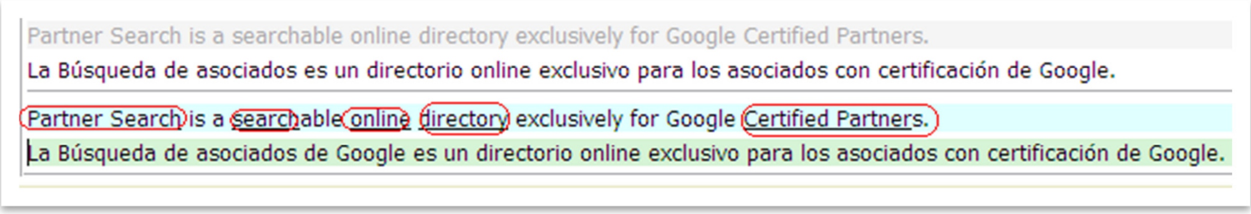
The KnifeCenter is looking for more information to give you, and information al

The KnifeCenter is constantly looking for more information to convey to you, .
El KnifeCenter está en constante búsqueda de la información que le pueda a

85% Fuzzy Match

Figure 14 - Fuzzy matches

Finally, term matches are retrieved thanks to the terminology database, which is usually integrated in the translation memory. The TM shows terminology matches to the translators that are stored in the terminology database.



Partner Search is a searchable online directory exclusively for Google Certified Partners.
La Búsqueda de asociados es un directorio online exclusivo para los asociados con certificación de Google.

Partner Search is a searchable online directory exclusively for Google Certified Partners.
La Búsqueda de asociados de Google es un directorio online exclusivo para los asociados con certificación de Google.

Figure 15 - Term matches

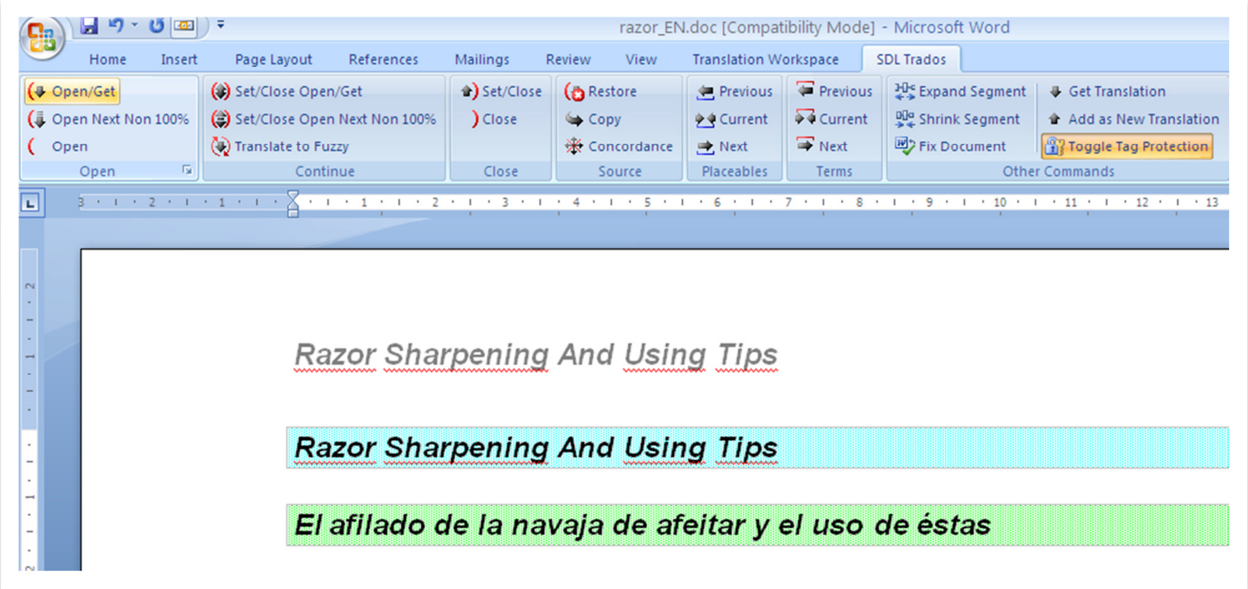
There are three main kinds of translation memory tools:

- Translation memory tools integrated in an application.

- Translation memory tools that are not integrated in an application.

- Translation memory tools that are especially designed for translation projects.

The first kind of TM tool is integrated with an editor program such as Microsoft Word, as *Trados Workbench* and *Wordfast*. Figure 16 shows a text processed with Trados Workbench.



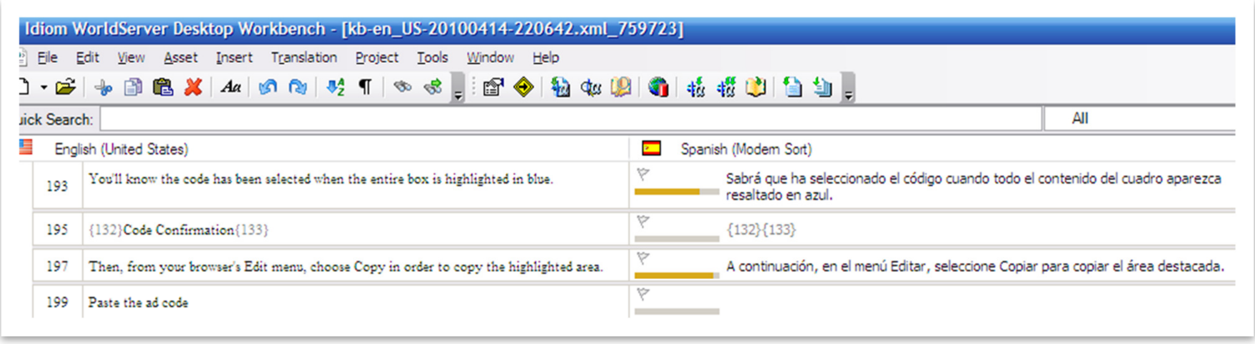
Razor Sharpening And Using Tips

Razor Sharpening And Using Tips

El afilado de la navaja de afeitar y el uso de éstas

Figure 16 - Trados Workbench

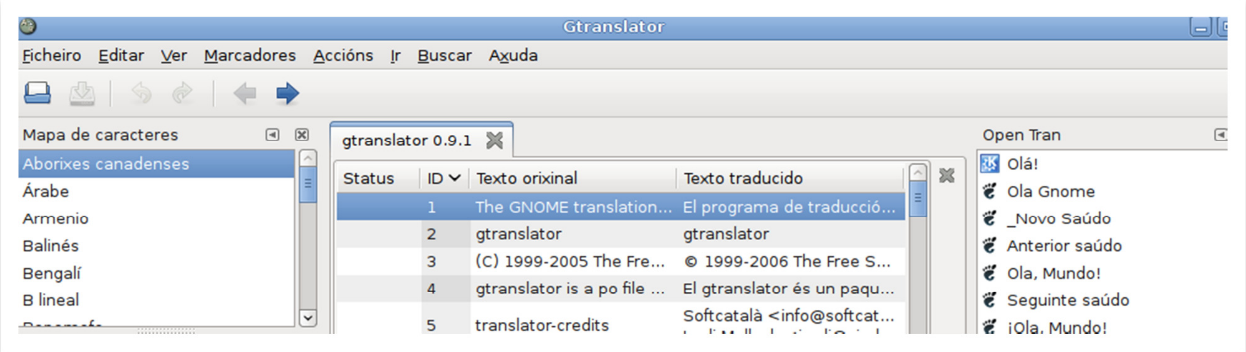
The second kind of TM tool is not integrated with an editor program; they import the document to their translation environment and once the translation is finished the document can be exported back to their original format. Some examples are Idiom WorldServer, Déjà Vu-X and SDL Trados Studio. Figure 17 is an example of a document imported into Idiom WorldServer.


Idiom WorldServer Desktop Workbench - [kb-en_US-20100414-220642.xml_759723]
File
Edit
View
Asset
Insert
Translation
Project
Tools
Window
Help
uick Search:
All

	English (United States)	Spanish (Modern Sort)
193	You'll know the code has been selected when the entire box is highlighted in blue.	Sabrá que ha seleccionado el código cuando todo el contenido del cuadro aparezca resaltado en azul.
195	{132}Code Confirmation{133}	{132}{133}
197	Then, from your browser's Edit menu, choose Copy in order to copy the highlighted area.	A continuación, en el menú Editar, seleccione Copiar para copiar el área destacada.
199	Paste the ad code	

Figure 17 - Idiom WorldServer

The third TM tool has been especially designed for translation projects. For example, Gtranslator is a TM tool that has been designed for translators who take part in free software translation projects. It was built for the translation of *Gnome Desktop* from Linux. (Oliver et al., 2007, pp. 30-35)



Status	ID	Texto orixinal	Texto traducido
	1	The GNOME translation...	El programa de traducció...
	2	gtranslator	gtranslator
	3	(C) 1999-2005 The Fre...	© 1999-2006 The Free S...
	4	gtranslator is a po file ...	El gtranslator és un paqu...
	5	translator-credits	Softcatalà <info@softcat...

Figure 18 - Gtranslator

As technology has imposed new skills to be acquired by translators, it is worth commenting briefly on translation training. Although most translators' knowledge is self-taught along their career, proper translation technology training is essential. According to Amparo Alcina (2008, p.96), translation technology training should cover five main areas:

1. The translators' computer equipment

2. Communication and documentation tools

3. Text edition and desktop publishing

4. Language tools and resources

5. Translation tools.

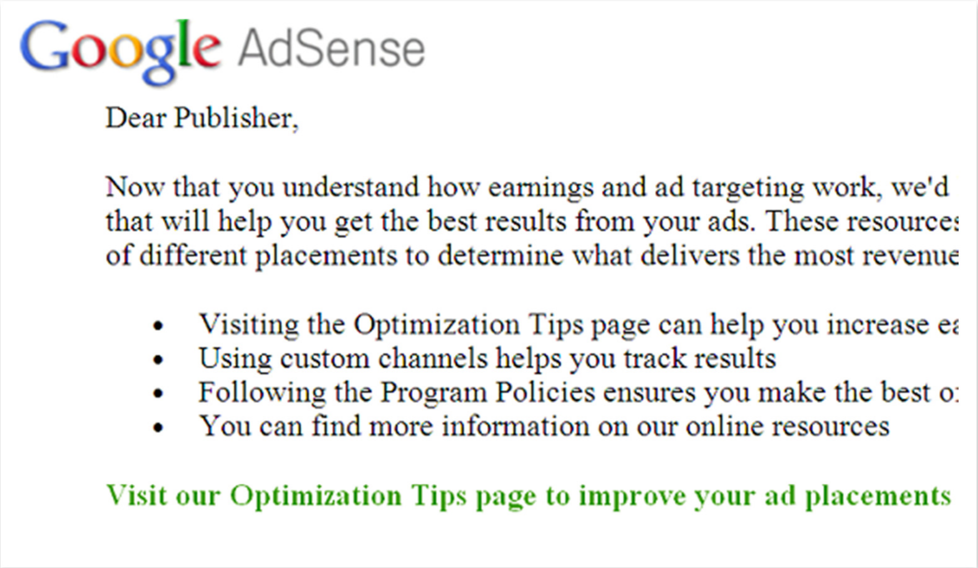
XML and other markup languages should also be included in translation trainings, although experts do not agree on the extent to which students should be familiar with XML, as Olga Nuñez (2006) affirms in her summary of an online discussion about how to teach XML. In that discussion, Mark Shuttleworth (Imperial College, London) states that trainings should contain an XML module that combines a theoretical component with practical work involving the creation of simple documents. This proposal seems realistic within the current translation studies framework.

3.6. TRANSLATION OF MARKUP LANGUAGE TEXTS

According to Bert Esselink (2000, pp. 213-219), there are four groups of tools that make possible the translation of markup language files: WYSIWYG editors, Text editors, Translation Memory tools and Protected tag editors.

3.6.1. WYSIWYG

WYSIWYG stands for "What You See Is What You Get". WYSIWYG HTML or XML editors are applications that enable the user to translate files where all page and text layout is visible and the markup is not displayed, for example, Microsoft Front Page. The main advantage of WYSIWYG editors is that translators can see the layout of the text, whereas the disadvantage is that the markup can be changed, deleted or added automatically if the HTML editor happens to re-write the code. Figure 18 below shows an HTML file opened to be translated with Microsoft Word.



Google AdSense

Dear Publisher,

Now that you understand how earnings and ad targeting work, we'd
that will help you get the best results from your ads. These resource
of different placements to determine what delivers the most revenue

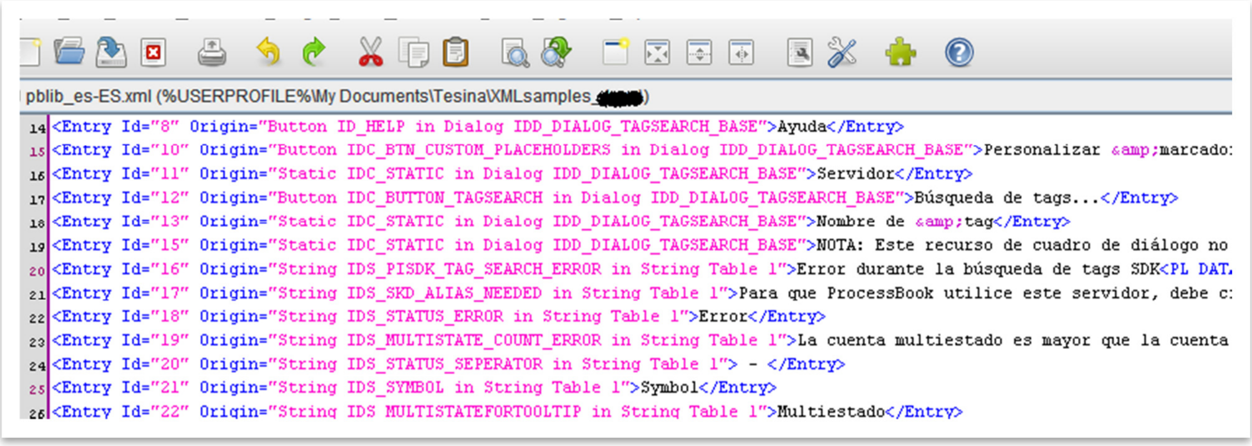
- Visiting the Optimization Tips page can help you increase e
- Using custom channels helps you track results
- Following the Program Policies ensures you make the best o
- You can find more information on our online resources

Visit our Optimization Tips page to improve your ad placements

Figure 19 - WYSIWYG

3.6.2. Text Editors

Text editors entail editing the source code directly. As a consequence, translators need to locate themselves the translatable text in the source text. The main disadvantages are that the layout of the HTML/XML file is not visible, that it is easy to corrupt the HTML code by deleting some tags, and that translators might translate text that should not be translated or even overlook translatable text.



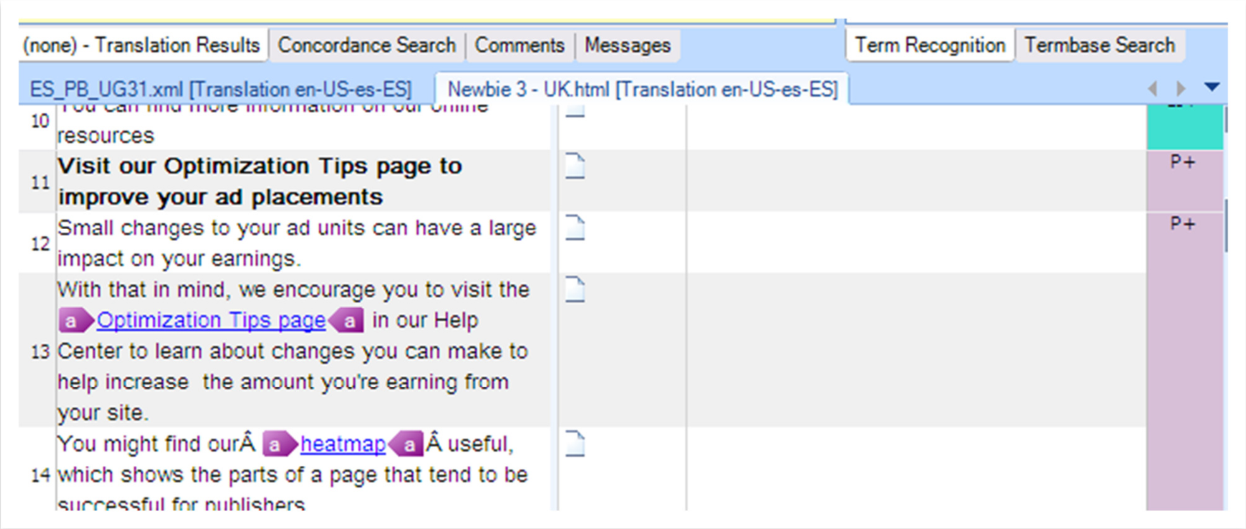
pblib_es-ES.xml (%USERPROFILE%\My Documents\Tesina\XMLsamples_[redacted])

```
<Entry Id="8" Origin="Button ID_HELP in Dialog IDD_DIALOG_TAGSEARCH_BASE">Ayuda</Entry>
<Entry Id="10" Origin="Button IDC_BTN_CUSTOM_PLACEHOLDERS in Dialog IDD_DIALOG_TAGSEARCH_BASE">Personalizar &amp;marcado:
<Entry Id="11" Origin="Static IDC_STATIC in Dialog IDD_DIALOG_TAGSEARCH_BASE">Servidor</Entry>
<Entry Id="12" Origin="Button IDC_BUTTON_TAGSEARCH in Dialog IDD_DIALOG_TAGSEARCH_BASE">Búsqueda de tags...</Entry>
<Entry Id="13" Origin="Static IDC_STATIC in Dialog IDD_DIALOG_TAGSEARCH_BASE">Nombre de &amp;tag</Entry>
<Entry Id="15" Origin="Static IDC_STATIC in Dialog IDD_DIALOG_TAGSEARCH_BASE">NOTA: Este recurso de cuadro de diálogo no
<Entry Id="16" Origin="String IDS_PISDK_TAG_SEARCH_ERROR in String Table 1">Error durante la búsqueda de tags SDK<PL DAT.
<Entry Id="17" Origin="String IDS_SKD_ALIAS_NEEDED in String Table 1">Para que ProcessBook utilice este servidor, debe c:
<Entry Id="18" Origin="String IDS_STATUS_ERROR in String Table 1">Error</Entry>
<Entry Id="19" Origin="String IDS_MULTISTATE_COUNT_ERROR in String Table 1">La cuenta multiestado es mayor que la cuenta
<Entry Id="20" Origin="String IDS_STATUS_SEPERATOR in String Table 1"> - </Entry>
<Entry Id="21" Origin="String IDS_SYMBOL in String Table 1">Symbol</Entry>
<Entry Id="22" Origin="String IDS_MULTISTATEFORTOOLTIP in String Table 1">Multiestado</Entry>
```

Figure 20 - Text editors

3.6.3. Translation Memory Tools

Translation memory tools, such as STAR Transit or SDL Trados Studio, contain filters for HTML or XML files. These tools either externalise all markup and only display the translatable text to the translator, or flag all markup using different colors and text styles.


(none) - Translation Results
Concordance Search
Comments
Messages
Term Recognition
Termbase Search
ES_PB_UG31.xml [Translation en-US-es-ES]
Newbie 3 - UK.html [Translation en-US-es-ES]
P+
P+

10 You can find more information on our online resources

11 **Visit our Optimization Tips page to improve your ad placements**

12 Small changes to your ad units can have a large impact on your earnings.

13 With that in mind, we encourage you to visit the a Optimization Tips page a in our Help Center to learn about changes you can make to help increase the amount you're earning from your site.

14 You might find ourÂ a heatmap a Â useful, which shows the parts of a page that tend to be successful for publishers

Figure 21 - HTML in SDL Trados Studio

3.6.4. Tag Editors

Tag editors, which have an integrated TM, enable translators to work in the HTML or XML source file where only the relevant markup is visible to translators. The best known Tag Editor is Trados TagEditor (see figure 22). Tag Editors protect all tags in the file from being deleted or overwritten and allow users to specify markup text settings. The latter is a very useful functionality because it enables translators to select the view of the tags: no tag text, partial tag text and complete tag text. On figure 22, the partial tag text view is activated.

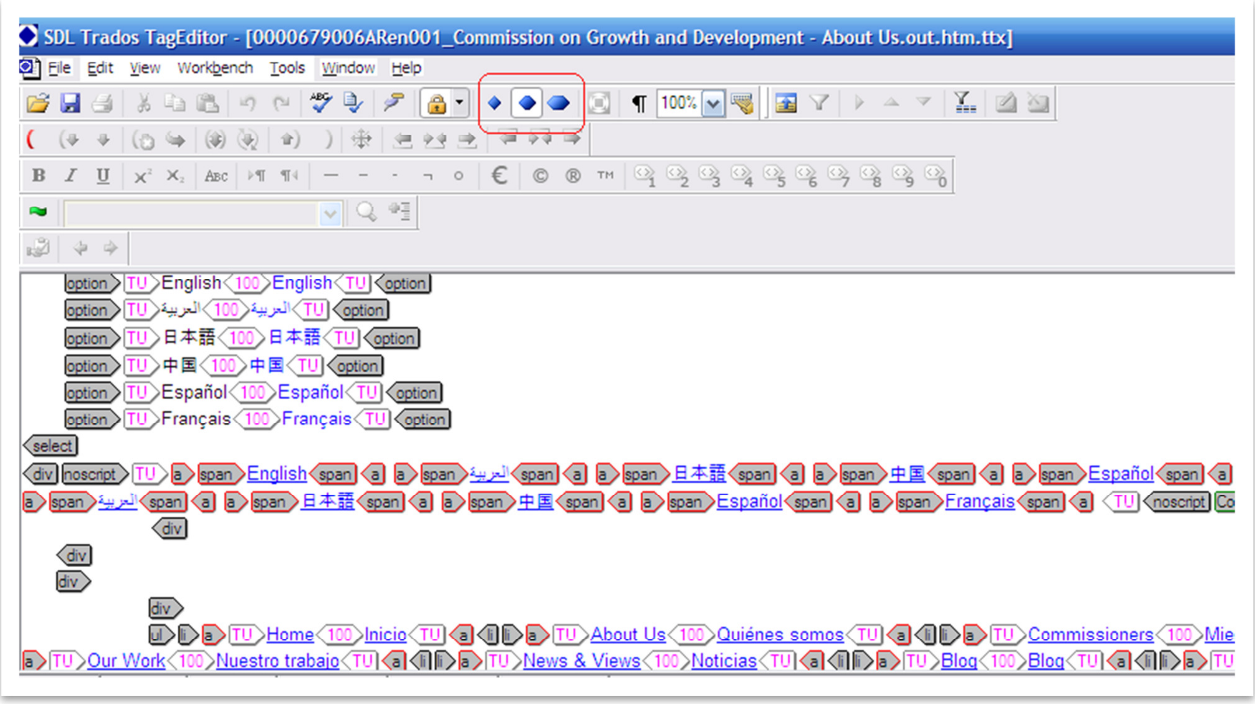

SDL Trados TagEditor - [0000679006ARen001_Commission on Growth and Development - About Us.out.htm.ttx]
File Edit View Workbench Tools Window Help
100%
option TU English 100 English TU option
option TU العربية 100 العربية TU option
option TU 日本語 100 日本語 TU option
option TU 中国 100 中国 TU option
option TU Español 100 Español TU option
option TU Français 100 Français TU option
select
div noscript TU a span English span a a span العربية span a a span 日本語 span a a span 中国 span a a span Español span a
a span العربية span a a span 日本語 span a a span 中国 span a a span Español span a a span Français span a TU noscript Co
div
div
div
div
ul li a TU Home 100 Inicio TU a li li a TU About Us 100 Quiénes somos TU a li li a TU Commissioners 100 Mie
a TU Our Work 100 Nuestro trabajo TU a li li a TU News & Views 100 Noticias TU a li li a TU Blog 100 Blog TU a li li a TU

Figure 22 - Trados TagEditor

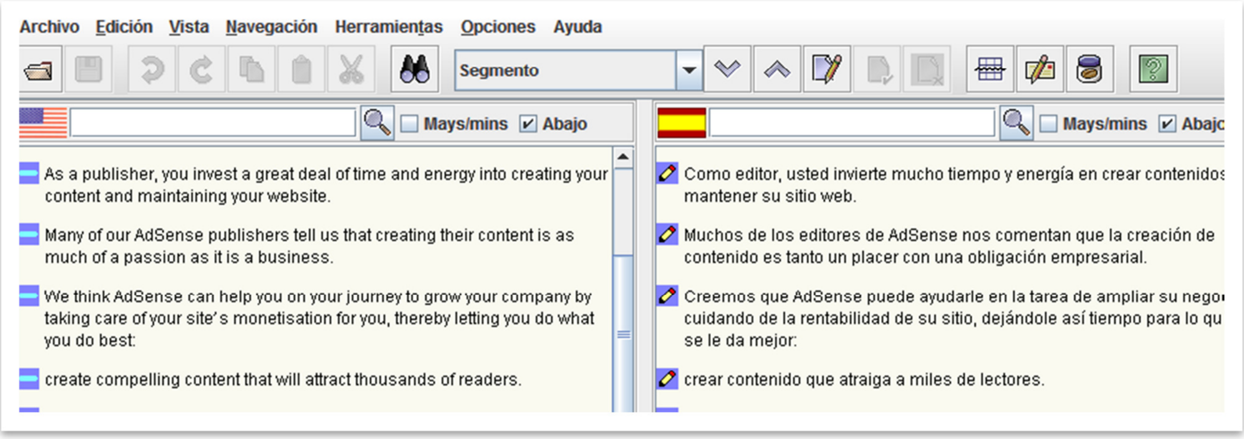
3.7. XLIFF editors

Since this research project is concerned with how XLIFF can affect the translation process, it is relevant to comment on the tools that are used to translate XLIFF files, as they have a direct impact on the translation process and quality.

There are two main sets of tools used to translate XLIFF files: XLIFF translation editors and XML editors.

1. XLIFF translation editors are integrated with translation memories and terminology tools. There are two main kinds of translation tools to translate XLIFF files:

   a) Specific tools designed to translate XLIFF files, such as Open Language Tools XLIFF Translation editor.



Archivo Edición Vista Navegación Herramientas Opciones Ayuda
Segmento
Mays/mins
Abajo
Mays/mins
Abajo

As a publisher, you invest a great deal of time and energy into creating your content and maintaining your website.

Many of our AdSense publishers tell us that creating their content is as much of a passion as it is a business.

We think AdSense can help you on your journey to grow your company by taking care of your site's monetisation for you, thereby letting you do what you do best:

create compelling content that will attract thousands of readers.

Como editor, usted invierte mucho tiempo y energía en crear contenidos mantener su sitio web.

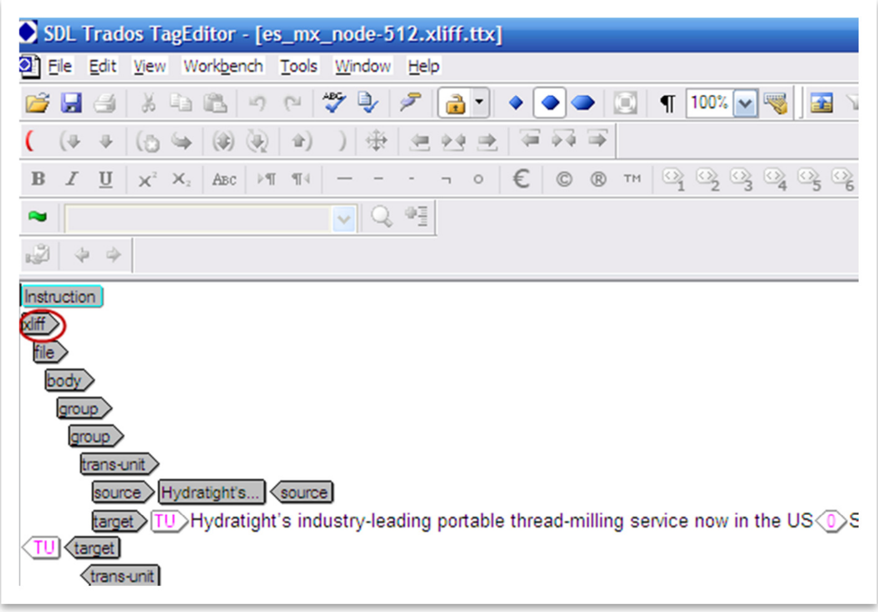
Muchos de los editores de AdSense nos comentan que la creación de contenido es tanto un placer con una obligación empresarial.

Creemos que AdSense puede ayudarle en la tarea de ampliar su nego cuidando de la rentabilidad de su sitio, dejándole así tiempo para lo qu se le da mejor:

crear contenido que atraiga a miles de lectores.

Figure 23 - Open Language Tools XLIFF Translation Editor 1

b) Translation memory tools that have filters for XLIFF files, such as Trados TagEditor. As XLIFF is based on XML, most translation memory tools that have an XML filter will allow the use of XLIFF.

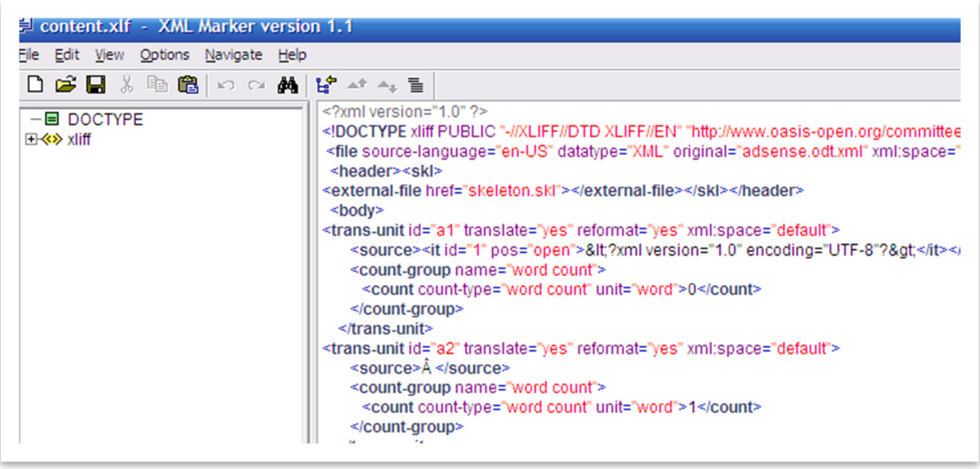


Hydratight's industry-leading portable thread-milling service now in the US

Figure 24 - TagEditor with XLIFF

2. XML or Text editors. As XLIFF is based on XML, XML editors allow the view and translation of XLIFF files. They are basically text editors that have some specific features to manage XML tags, such a protection to avoid overwriting the code. XML text editors also have functionalities for developers.

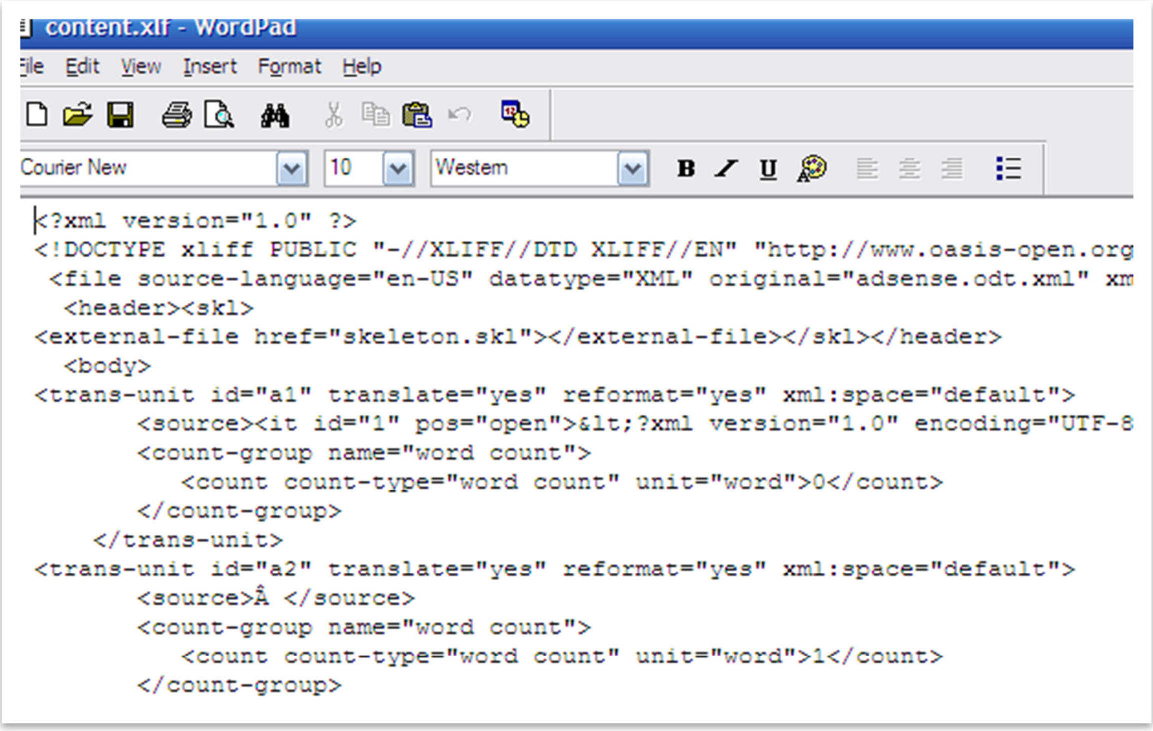
See below a screenshot of XML Marker:



```
<?xml version="1.0" ?>
<!DOCTYPE xliff PUBLIC "-//XLIFF//DTD XLIFF//EN" "http://www.oasis-open.org/committee
 <file source-language="en-US" datatype="XML" original="adsense.odt.xml" xml:space="
  <header><skl>
<external-file href="skeleton.skl"></external-file></skl></header>
  <body>
<trans-unit id="a1" translate="yes" reformat="yes" xml:space="default">
      <source><it id="1" pos="open">&lt;?xml version="1.0" encoding="UTF-8"?&gt;</it></
      <count-group name="word count">
        <count count-type="word count" unit="word">0</count>
      </count-group>
    </trans-unit>
<trans-unit id="a2" translate="yes" reformat="yes" xml:space="default">
      <source>Â </source>
      <count-group name="word count">
        <count count-type="word count" unit="word">1</count>
      </count-group>
```

Figure 25 - XML marker

These XML or text editors use different colours to identify the various markup elements easily. If a plain text editor is used to translate an XLIFF file, the complex view of the document can cause great confusion to translators, as it can be quite difficult to separate translatable text from non-translatable text. Figure 26 shows a screenshot from the same XLIFF document as figure 25 but opened with the plain text editor Wordpad.



```
<?xml version="1.0" ?>
<!DOCTYPE xliff PUBLIC "-//XLIFF//DTD XLIFF//EN" "http://www.oasis-open.org
 <file source-language="en-US" datatype="XML" original="adsense.odt.xml" xm
  <header><skl>
<external-file href="skeleton.skl"></external-file></skl></header>
  <body>
<trans-unit id="a1" translate="yes" reformat="yes" xml:space="default">
       <source><it id="1" pos="open">&lt;?xml version="1.0" encoding="UTF-8
       <count-group name="word count">
          <count count-type="word count" unit="word">0</count>
       </count-group>
    </trans-unit>
<trans-unit id="a2" translate="yes" reformat="yes" xml:space="default">
       <source>Â </source>
       <count-group name="word count">
          <count count-type="word count" unit="word">1</count>
       </count-group>
```

Figure 26 - WordPad

Text editors are not integrated with translation memories, which is a major disadvantage for the translator. Consequently, it is not so common anymore to use text editors for translation.

These tools and the way they interact with XLIFF are analysed in more detail in the “Case Study” section, as they are used to prove the hypothesis presented in this research project.

4. CASE STUDY

4.1. Methodology

As it has been exposed in the “Aim and Scope of the Research” section, the aim of this research project is:

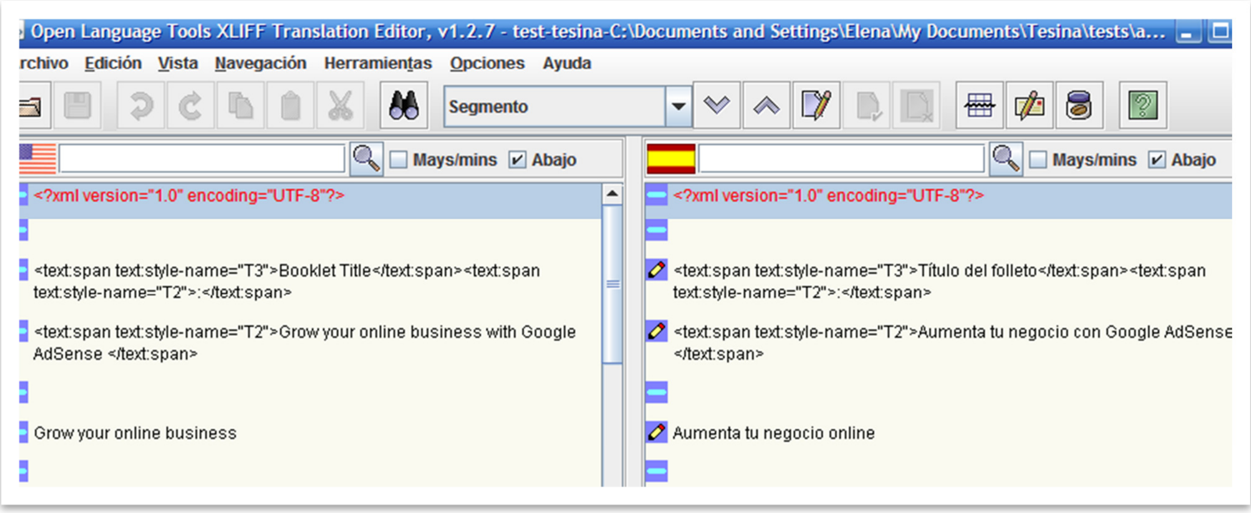
a) To determine whether the XLIFF format adds any value or hinders the translation process.

b) To find out if XLIFF editor tools hide any relevant information for translators.

c) To provide a proposal of how information should be marked by XLIFF editors.

In order to carry out this research, XLIFF files have been observed in two ways: a) with three different XLIFF editors and b) with XML or text editors.

a) Viewed with XLIFF editors


Open Language Tools XLIFF Translation Editor, v1.2.7 - test-tesina-C:\Documents and Settings\Elena\My Documents\Tesina\tests\a...
rchivo
Edición
Vista
Navegación
Herramientas
Opciones
Ayuda
Segmento
□ Mays/mins
☑ Abajo
□ Mays/mins
☑ Abajo

<?xml version="1.0" encoding="UTF-8"?>	<?xml version="1.0" encoding="UTF-8"?>
<text:span text:style-name="T3">Booklet Title</text:span><text:span text:style-name="T2">:</text:span>	<text:span text:style-name="T3">Título del folleto</text:span><text:span text:style-name="T2">:</text:span>
<text:span text:style-name="T2">Grow your online business with Google AdSense </text:span>	<text:span text:style-name="T2">Aumenta tu negocio con Google AdSense </text:span>
Grow your online business	Aumenta tu negocio online

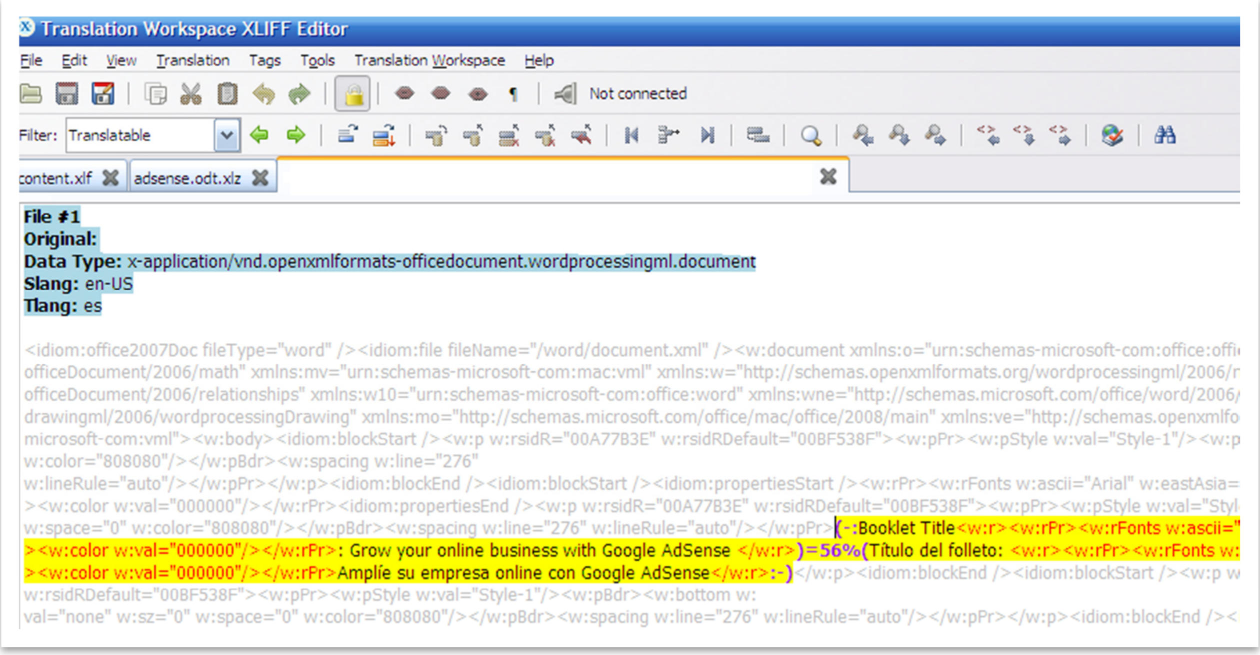
Figure 27 - Open Language Tools XLIFF Translation Editor 2

XLIFF editors filter XLIFF information and often hide tags (or have the option of hiding them) in order to show translators only translatable text.

In this research project the following XLIFF editors have been used: Open Language Tools XLIFF Translation Editor, Translation Workspace XLIFF Editor and SDL Trados Studio 2009. All of these are translation memory tools.

Open Language Tools are a set of translation tools that aim at making the task of translating software and documentation easier. They are comprised of a full-featured XLIFF Translation Editor and a set of XLIFF file-filters for a number of documentation and software file formats. This program has been chosen mainly because it is free software[6] and because it can be interesting to compare it with commercial software.

Translation Workspace is a cloud computing[7] CAT tool, where the technology has been created by Lionbridge Technologies. Up to a few months ago, it used to be called Logoport though now it is being commercialised as Translation Workspace. Lionbridge is a leading provider of language, development and testing services. The tool has been incorporated in this project because it is used by a vast number of freelance translators.


Translation Workspace XLIFF Editor
File Edit View Translation Tags Tools Translation Workspace Help
Not connected
Filter: Translatable
content.xlf
adsense.odt.xlz

File #1
Original:
Data Type: x-application/vnd.openxmlformats-officedocument.wordprocessingml.document
Slang: en-US
Tlang: es

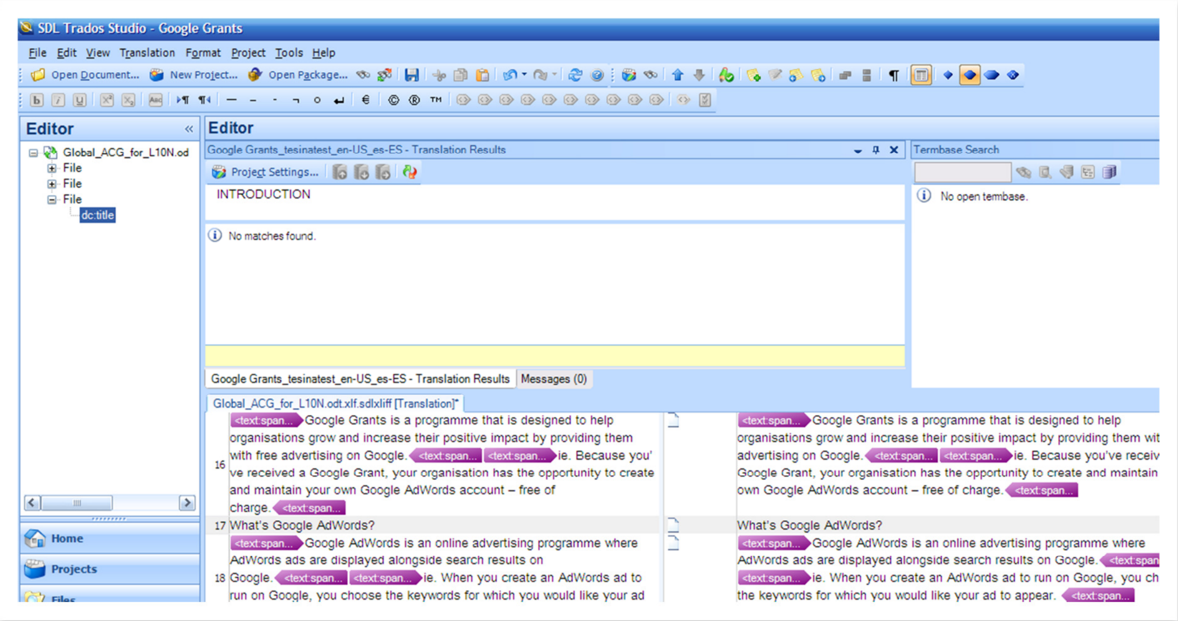
```
<idiom:office2007Doc fileType="word" /><idiom:file fileName="/word/document.xml" /><w:document xmlns:o="urn:schemas-microsoft-com:office:offi
officeDocument/2006/math" xmlns:mv="urn:schemas-microsoft-com:mac:vml" xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/r
officeDocument/2006/relationships" xmlns:w10="urn:schemas-microsoft-com:office:word" xmlns:wne="http://schemas.microsoft.com/office/word/2006/
drawingml/2006/wordprocessingDrawing" xmlns:mo="http://schemas.microsoft.com/office/mac/office/2008/main" xmlns:ve="http://schemas.openxmlfo
microsoft-com:vml"><w:body><idiom:blockStart /><w:p w:rsidR="00A77B3E" w:rsidRDefault="00BF538F"><w:pPr><w:pStyle w:val="Style-1"/><w:p
w:color="808080"/></w:pBdr><w:spacing w:line="276"
w:lineRule="auto"/></w:pPr></w:p><idiom:blockEnd /><idiom:blockStart /><idiom:propertiesStart /><w:rPr><w:rFonts w:ascii="Arial" w:eastAsia=
><w:color w:val="000000"/></w:rPr><idiom:propertiesEnd /><w:p w:rsidR="00A77B3E" w:rsidRDefault="00BF538F"><w:pPr><w:pStyle w:val="Styl
w:space="0" w:color="808080"/></w:pBdr><w:spacing w:line="276" w:lineRule="auto"/></w:pPr>(-:Booklet Title<w:r><w:rPr><w:rFonts w:ascii="
><w:color w:val="000000"/></w:rPr>: Grow your online business with Google AdSense </w:r>)=56%(Título del folleto: <w:r><w:rPr><w:rFonts w:
><w:color w:val="000000"/></w:rPr>Amplíe su empresa online con Google AdSense</w:r>:-)</w:p><idiom:blockEnd /><idiom:blockStart /><w:p w
w:rsidRDefault="00BF538F"><w:pPr><w:pStyle w:val="Style-1"/><w:pBdr><w:bottom w:
val="none" w:sz="0" w:space="0" w:color="808080"/></w:pBdr><w:spacing w:line="276" w:lineRule="auto"/></w:pPr></w:p><idiom:blockEnd /><
```

Figure 28- Translation Workspace XLIFF editor

Finally, SDL Trados Studio 2009 is an application for translating files, creating and managing translation memories. This product has been developed by SDL Trados, which is the market leading computer assisted translation software suites. This tool has been used in the Case Study

[6] Free software is software that gives the user the freedom to share, study and modify it. (FSF, 2010)
[7] Cloud computing is Internet-based application, whereby shared resources, software, and information are provided to computers and other devices on demand.

because its latest version, SDL Trados Studio 2009, allows translating XLIFF files, and because Trados is one of the most important translation tools. SDL Trados must be acquired through a license.


SDL Trados Studio - Google Grants
File Edit View Translation Format Project Tools Help
Open Document... New Project... Open Package...
Editor
Global_ACG_for_L10N.od
File
File
File
dc:title
Home
Projects
Files
Editor
Google Grants_tesinatest_en-US_es-ES - Translation Results
Project Settings...
Termbase Search
Google Grants_tesinatest_en-US_es-ES - Translation Results
Messages (0)
Global_ACG_for_L10N.odt.xlf.sdlxliff [Translation]*

INTRODUCTION

No matches found.

No open termbase.

16 <text:span...Google Grants is a programme that is designed to help organisations grow and increase their positive impact by providing them with free advertising on Google.<text:span... <text:span...ie. Because you've received a Google Grant, your organisation has the opportunity to create and maintain your own Google AdWords account – free of charge.<text:span...

<text:span...Google Grants is a programme that is designed to help organisations grow and increase their positive impact by providing them wit advertising on Google.<text:span... <text:span...ie. Because you've receiv Google Grant, your organisation has the opportunity to create and maintain own Google AdWords account – free of charge.<text:span...

17 What's Google AdWords?

What's Google AdWords?

18 <text:span...Google AdWords is an online advertising programme where AdWords ads are displayed alongside search results on Google.<text:span... <text:span...ie. When you create an AdWords ad to run on Google, you choose the keywords for which you would like your ad

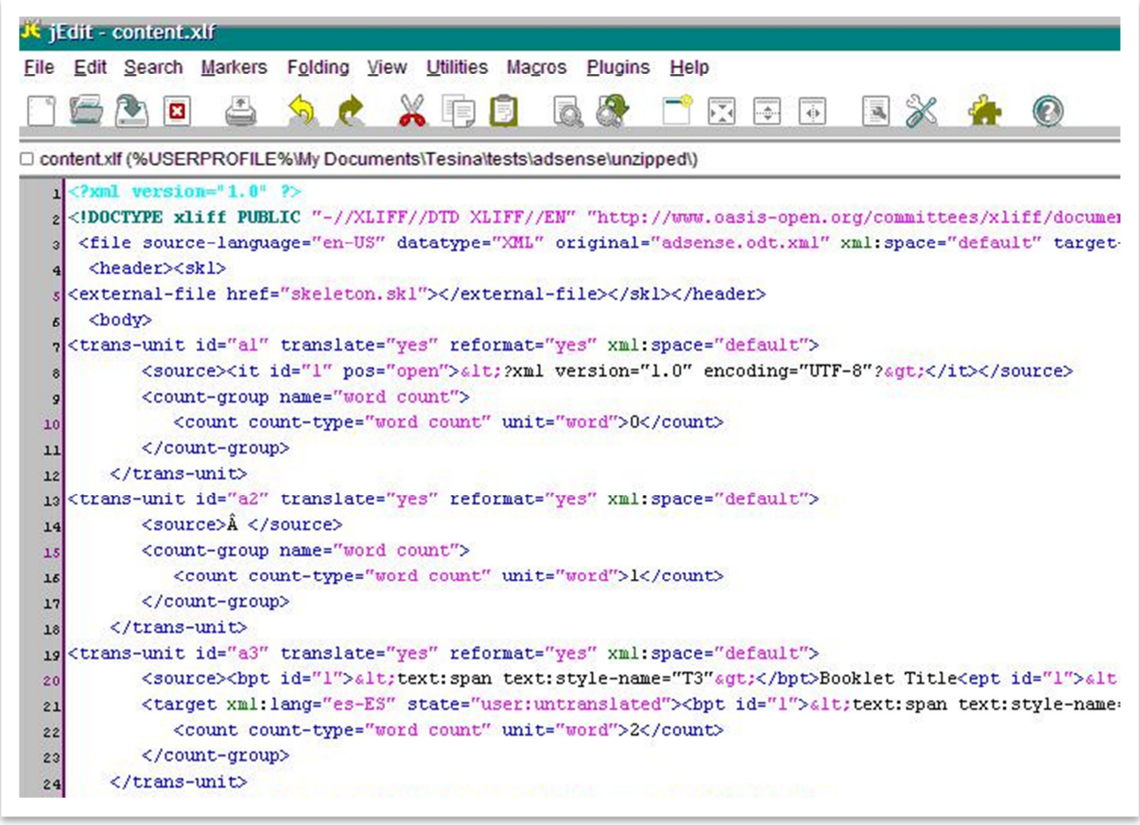
<text:span...Google AdWords is an online advertising programme where AdWords ads are displayed alongside search results on Google.<text:span <text:span...ie. When you create an AdWords ad to run on Google, you ch the keywords for which you would like your ad to appear. <text:span...

Figure 29 - SDL Trados Studio

b) Viewed with an XML editor

As explained before in the "Translation Tools for Markup Languages" section, XML editors are text editors that will display the files as they are, and will show all tags without any kind of filtering, thus having access to the “raw file”. However, it is important to clarify that XLIFF files are typically not source files, but converted or filtered files from a source text (HTML, DOC, ODT, etc). This is important because file converters use an implementation of XLIFF in order to convert the source file into XLIFF, therefore the converted XLIFF file might not benefit from all XLIFF specifications defined by OASIS. This implementation of XLIFF is designed by developers.

The XML editor chosen for this project is jEdit, which is a programmer´s text editor and free software (see figure 30 below). In some occasions, XML Marker has been used, which is also a free software text editor.


jEdit - content.xlf
File Edit Search Markers Folding View Utilities Macros Plugins Help

□ content.xlf (%USERPROFILE%\My Documents\Tesina\tests\adsense\unzipped\)

```
<?xml version="1.0" ?>
<!DOCTYPE xliff PUBLIC "-//XLIFF//DTD XLIFF//EN" "http://www.oasis-open.org/committees/xliff/documen
 <file source-language="en-US" datatype="XML" original="adsense.odt.xml" xml:space="default" target-
  <header><skl>
<external-file href="skeleton.skl"></external-file></skl></header>
  <body>
<trans-unit id="a1" translate="yes" reformat="yes" xml:space="default">
       <source><it id="1" pos="open">&lt;?xml version="1.0" encoding="UTF-8"?&gt;</it></source>
       <count-group name="word count">
          <count count-type="word count" unit="word">0</count>
       </count-group>
    </trans-unit>
<trans-unit id="a2" translate="yes" reformat="yes" xml:space="default">
       <source>Â </source>
       <count-group name="word count">
          <count count-type="word count" unit="word">1</count>
       </count-group>
    </trans-unit>
<trans-unit id="a3" translate="yes" reformat="yes" xml:space="default">
       <source><bpt id="1">&lt;text:span text:style-name="T3"&gt;</bpt>Booklet Title<ept id="1">&lt
       <target xml:lang="es-ES" state="user:untranslated"><bpt id="1">&lt;text:span text:style-name:
          <count count-type="word count" unit="word">2</count>
       </count-group>
    </trans-unit>
```

Figure 30 - XML editor

In terms of the corpora used for this research project, rich-layout documentation files have been collected to view how XLIFF treats the different aspects of the layout of the source text and how they are displayed by the XLIFF editors. Around twenty five XLIFF files have been observed, although not all of them have been used for the "Case Study". The word count of these files amounts to 44,000 words. This corpora volume is representative for the purposes of this research project. Most of the files are XLIFF texts provided by translators, and other texts are manuals chosen because of their rich-layout. It is important to point out that many of the samples analysed have not been used as examples in the Case Study because of the confidentially agreements signed by the translators who provided the texts. The corpora were comprised in the Appendix

A[8], although not all samples are included in the present published study due to the mentioned confidentiality issues.

The source file format of the samples used for this research is OpenDocument (.odt file), as it is usually compatible with XLIFF converters. It has been necessary to convert the OpenDocument files into XLIFF (.xlf files) in order to evaluate how XLIFF processes the layout of the source text. Basically, this conversion has been performed by using two tools: Open Language Tools XLIFF filters, which turns a variety of file formats into XLIFF files, and Okapi´s Rainbow[9], which is a graphical user interface that allows users to specify some of the options of the utilities, such as a converter into XLIFF.

Once the files were converted into XLIFF, they were ready to be observed according to the two processes explained before (with an XLIFF editor and an XML editor).

The approach taken for the evaluation of XLIFF files was to observe the files as translators would find them for translation. Since there are many aspects that could be the object of this evaluation, a limited list of elements has been the focus of the case study:

- original file information

- XLM declaration

- headings

- images and captions

- lists

- footnotes

- links.

[8] Appendixes have been removed from this version for publication due to confidentiality issues.

[9] The Okapi Framework is a set of interface specifications, format definitions, components and applications that provides an environment to build interoperable tools for the different steps of the translation and localisation process.

Once these elements have been analysed, a proposal has been formulated about how XLIFF editors should display layout and information so that they are more easily identifiable for the translator. This proposal does not intend to provide a description on how improvements should be made from a technical point of view.

Finally, to better understand what translators think about XLIFF and what kind of information they miss while translating, a questionnaire has been distributed among professional translators. These were the questions submitted to them:

1. What programs do you use for the translation of documentation/online help files?
2. What kind of information do these programs allow you to see?
3. When you translate documentation/online help files, what is the most usual file format you get?
4. If you translate XLIFF files, which XLIFF editor do you use? Do you think XLIFF editors allow you to see more information than other applications about the text or the project?
5. When you translate documentation/online help files, what kind of information do you usually miss?

The translation mass has been reached through the following online translation groups:

- Elebe Freelance Translators: a group of freelance translators from Spain that work for Lionbridge Technologies.

- Traducción y Tecnologías: a group of translators that took the Postgraduate course Translation and Technology in the Universtitat Oberta de Catalunya (UOC).

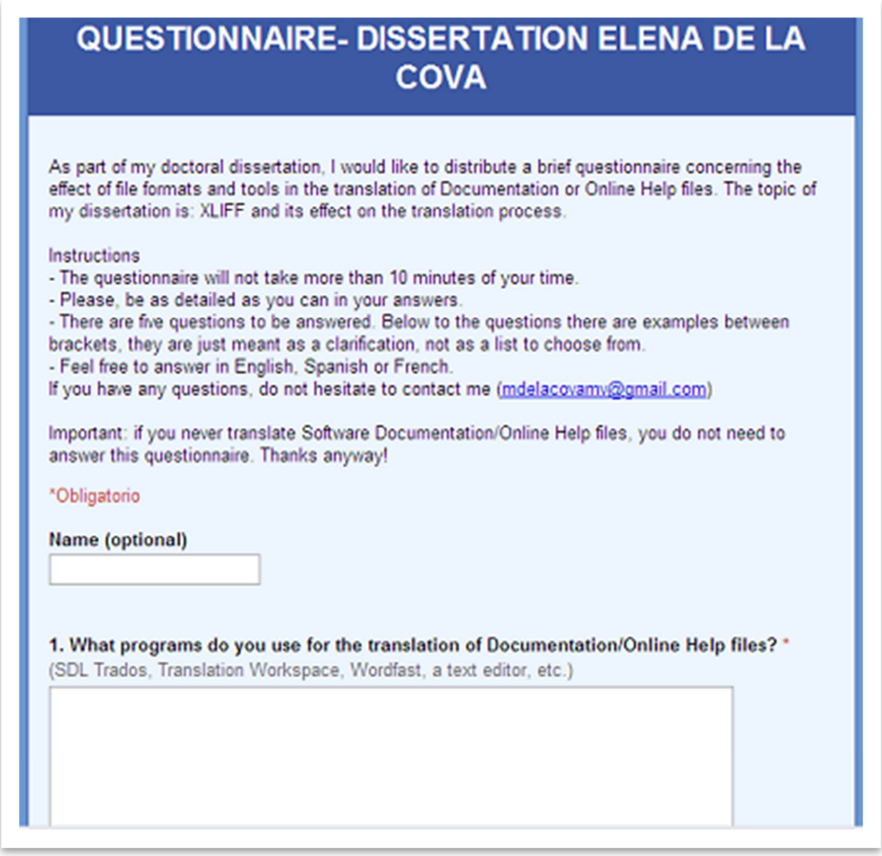
- Traducción España: a distribution list about translation in Spain which deals with various translation topics: translation tools, terminology, resources, work opportunities, translation business.

- ATD (Ágora de Traductors Digitals): a Catalan distribution list for translators that work with New Technologies.

- APTIC (Associació Professional de Traductors i Interprets de Catalunya): an independent, non-profit association.

In addition, the questionnaire has also been sent to a number of translators outside of the groups listed above, who in turn have circulated the questionnaire among other translators.

The questionnaire was sent out through a Google Form:



QUESTIONNAIRE- DISSERTATION ELENA DE LA COVA

As part of my doctoral dissertation, I would like to distribute a brief questionnaire concerning the effect of file formats and tools in the translation of Documentation or Online Help files. The topic of my dissertation is: XLIFF and its effect on the translation process.

Instructions
- The questionnaire will not take more than 10 minutes of your time.
- Please, be as detailed as you can in your answers.
- There are five questions to be answered. Below to the questions there are examples between brackets, they are just meant as a clarification, not as a list to choose from.
- Feel free to answer in English, Spanish or French.
If you have any questions, do not hesitate to contact me (mdelacovamv@gmail.com)

Important: if you never translate Software Documentation/Online Help files, you do not need to answer this questionnaire. Thanks anyway!

*Obligatorio

Name (optional)

1. What programs do you use for the translation of Documentation/Online Help files? *
(SDL Trados, Translation Workspace, Wordfast, a text editor, etc.)

Figure 31 - GoogleForm questionnaire

4.2. XLIFF analysis

The XLIFF corpora has been analysed according to the following criteria:

- original file information

- XML declaration

- headings

- images and captions

- lists

- footnotes

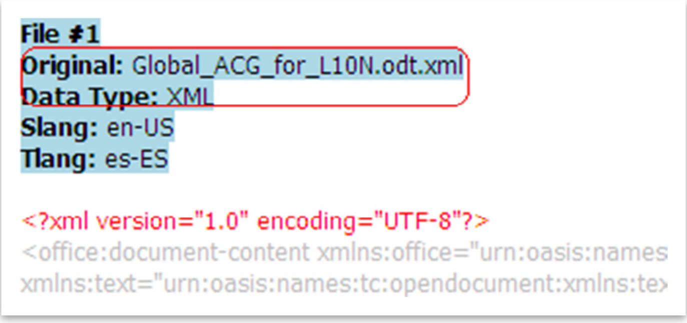
- links.

As mentioned before, the documents used for this study are displayed by three XLIFF tools and by a text editor. The samples have been chosen taking into account the richness of the layout, so that as much information as possible could be viewed in the files.

4.2.1. Original file

4.2.1.1. XLIFF editors

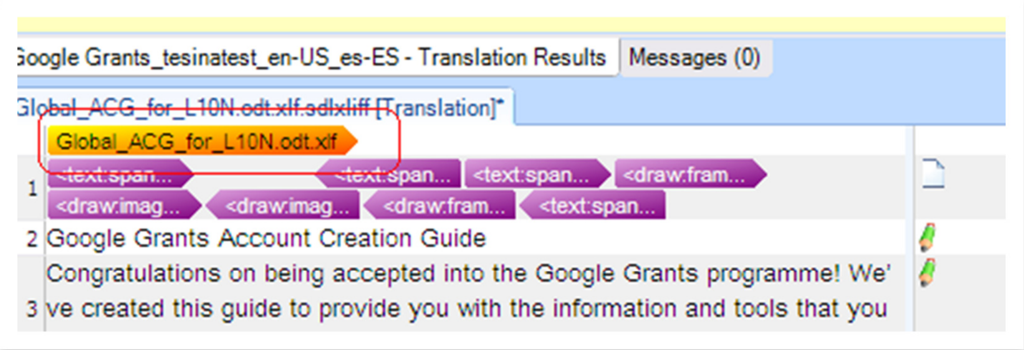
When opening an XLIFF file with the three XLIFF editors, it is found that they display different information. SDL Trados Studio 2009 and Translation Workspace show the name of the original file (figures 32 and 33). By contrast, Open Language Tools does not show the original file information.



File #1
Original: Global_ACG_for_L10N.odt.xml
Data Type: XML
Slang: en-US
Tlang: es-ES

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content xmlns:office="urn:oasis:names
xmlns:text="urn:oasis:names:tc:opendocument:xmlns:te>
```

Figure 32 - Source file in Translation Workspace


ɜoogle Grants_tesinatest_en-US_es-ES - Translation Results
Messages (0)
Global_ACG_for_L10N.odt.xlf.sdlxliff [Translation]*
Global_ACG_for_L10N.odt.xlf
<text:span...
<text:span...
<text:span...
<draw:fram...
<draw:imag...
<draw:imag...
<draw:fram...
<text:span...

1

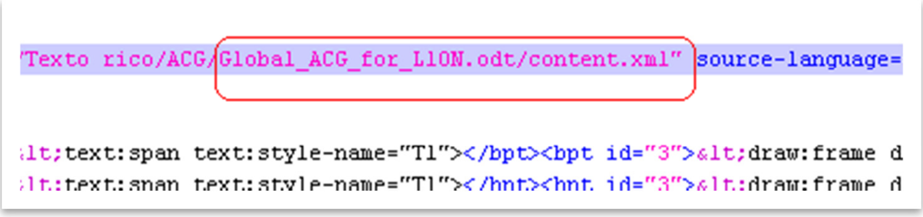
2 Google Grants Account Creation Guide

3 Congratulations on being accepted into the Google Grants programme! We've created this guide to provide you with the information and tools that you

Figure 33 - Source file in SDL Trados Studio

4.2.1.2. Text editor

The same XLIFF file is opened with the text editor jEdit and, as figure 34 shows, the source file name is also displayed in the document.



```
'Texto rico/ACG/Global_ACG_for_L10N.odt/content.xml" source-language=

lt;text:span text:style-name="T1"></bpt><bpt id="3">&lt;draw:frame d
lt;text:span text:style-name="T1"></bpt><bpt id="3">&lt;draw:frame d
```

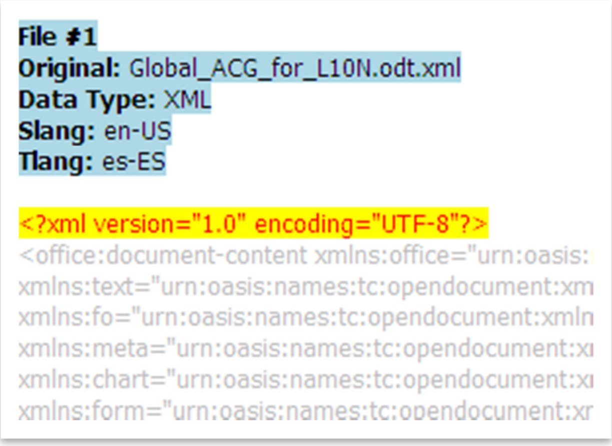
Figure 34 - Source file in jEdit

It is always relevant to see the name of the file, which provides information about the file extension and probably about the topic of the translation.

4.2.2. XML declaration

4.2.2.1. XLIFF editors

Only Open Language Tools and Translation Workspace show the XML declaration of the XLIFF file. The XLIFF format is based on XML and therefore the XML declaration must be the first line of the file. Note that the XML declaration is composed of the XML version and the encoding used in the document. Figure 35 shows the XML declaration of a document in the Translation Workspace XLIFF Editor.



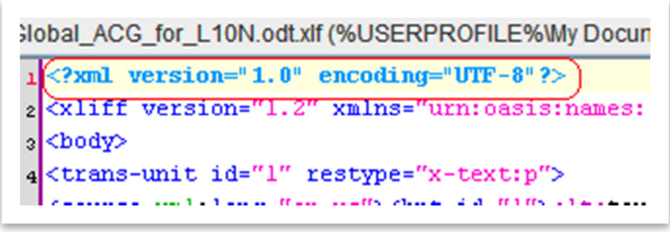
File #1
Original: Global_ACG_for_L10N.odt.xml
Data Type: XML
Slang: en-US
Tlang: es-ES

```
<?xml version="1.0" encoding="UTF-8"?>
<office:document-content xmlns:office="urn:oasis:
xmlns:text="urn:oasis:names:tc:opendocument:xm
xmlns:fo="urn:oasis:names:tc:opendocument:xmln
xmlns:meta="urn:oasis:names:tc:opendocument:x
xmlns:chart="urn:oasis:names:tc:opendocument:x
xmlns:form="urn:oasis:names:tc:opendocument:xr
```

Figure 35 - XML declaration in Translation Workspace

4.2.2.2. Text editors

The text editor jEdit also displays the XML declaration at the beginning of the document:



lobal_ACG_for_L10N.odt.xlf (%USERPROFILE%\My Docun

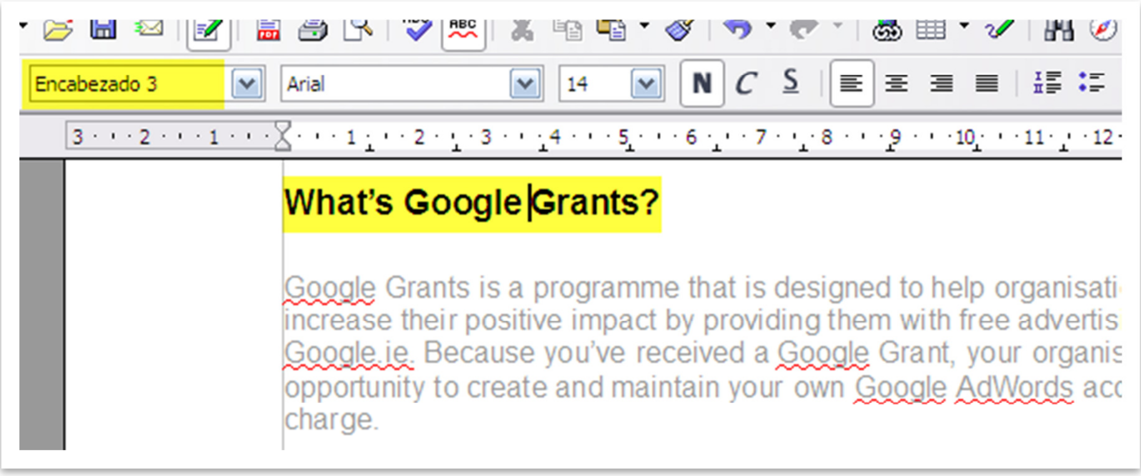
```
<?xml version="1.0" encoding="UTF-8"?>
<xliff version="1.2" xmlns="urn:oasis:names:
<body>
<trans-unit id="1" restype="x-text:p">
```

Figure 36 - XML declaration in jEdit

4.2.3. Headings

4.2.3.1. XLIFF editors

The Open Language Tools XLIFF Translation Editor does not flag headings in the text, even when titles are formatted with heading style in the original OpenDocument file, as can be seen in the example below:

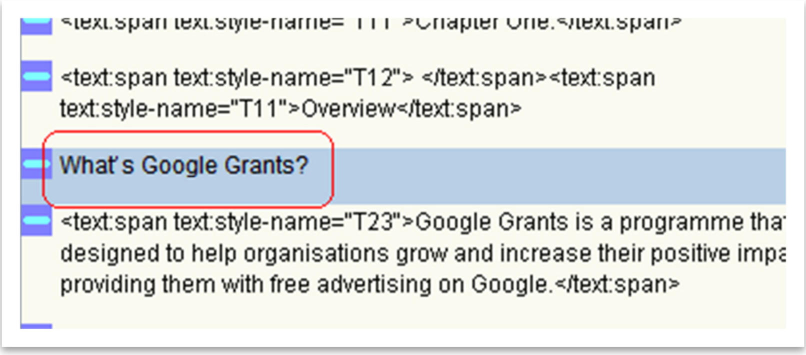


What's Google Grants?

Google Grants is a programme that is designed to help organisati
increase their positive impact by providing them with free advertis
Google.ie. Because you've received a Google Grant, your organis
opportunity to create and maintain your own Google AdWords acc
charge.

Figure 37 - Headings in source OpenDocument

This means that translators do not necessarily know if they are translating an isolated sentence from a paragraph, a heading or a subheading. Figure 38 shows the same example as figure 37 but converted into XLIFF and opened with Open Language Tools.

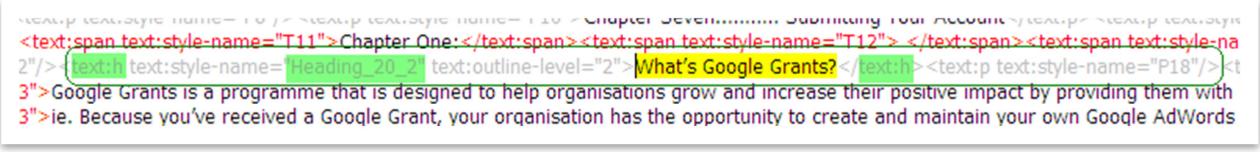


```
<text:span text:style-name="T11">Chapter One.</text:span>

<text:span text:style-name="T12"> </text:span><text:span
text:style-name="T11">Overview</text:span>

What's Google Grants?

<text:span text:style-name="T23">Google Grants is a programme tha
designed to help organisations grow and increase their positive impa
providing them with free advertising on Google.</text:span>
```

Figure 38 - Headings in Open Language Tools

Translation Workspace and SDL Trados Studio specify headings differently.

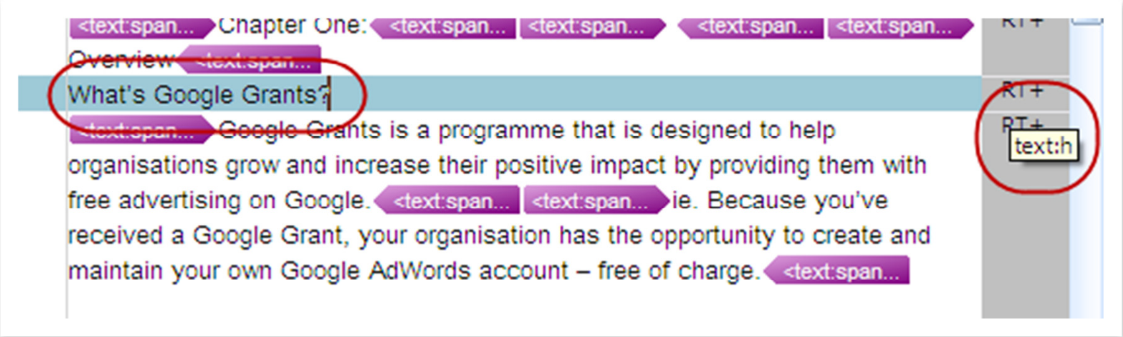
Translation Workspace shows the XLIFF tag that identifies the segment as a heading (<text:h>), as can be seen on the green highlighted elements from figure 39 (which corresponds to the source text from figure 37):



```
<text:p text:style-name="P8"/><text:p text:style-name="P10">Chapter Seven.......... Submitting Your Account</text:p><text:p text:style
<text:span text:style-name="T11">Chapter One:</text:span><text:span text:style-name="T12"> </text:span><text:span text:style-na
2"/><text:h text:style-name="Heading_20_2" text:outline-level="2">What's Google Grants?</text:h><text:p text:style-name="P18"/><t
3">Google Grants is a programme that is designed to help organisations grow and increase their positive impact by providing them with
3">ie. Because you've received a Google Grant, your organisation has the opportunity to create and maintain your own Google AdWords
```

Figure 39 - Headings in Translation Workspace

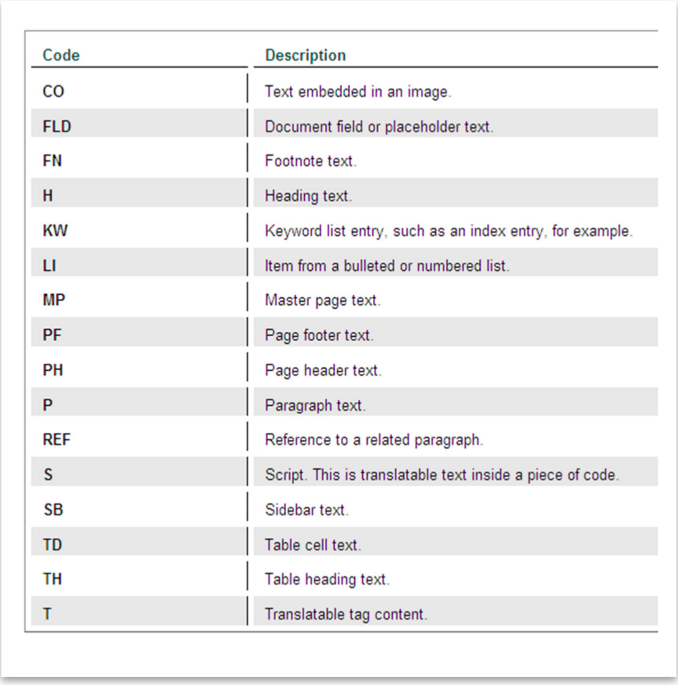
SDL Trados Studio does not show XLIFF tags as Translation Workspace does; however, it contains a column in the translation panel that highlights the document structure. Figure 40 also corresponds to the source text from figure 37.



<text:span... Chapter One: <text:span... <text:span... <text:span... <text:span...
Overview <text:span...
What's Google Grants?
<text:span... Google Grants is a programme that is designed to help organisations grow and increase their positive impact by providing them with free advertising on Google. <text:span... <text:span... ie. Because you've received a Google Grant, your organisation has the opportunity to create and maintain your own Google AdWords account – free of charge. <text:span...

Figure 40 - Headings in SDL Trados Studio

RT+ stands for Custom Resource Type, which refers to the location of the segment text in the source document. To see what kind of element each segment corresponds to, it is necessary to place the cursor over the column after which the cursor changes to a hand , which enables the user to see a code for the structure of the segment. In figure 40, "h" stands for "heading". See below an example of the typical codes for the structure of the document.

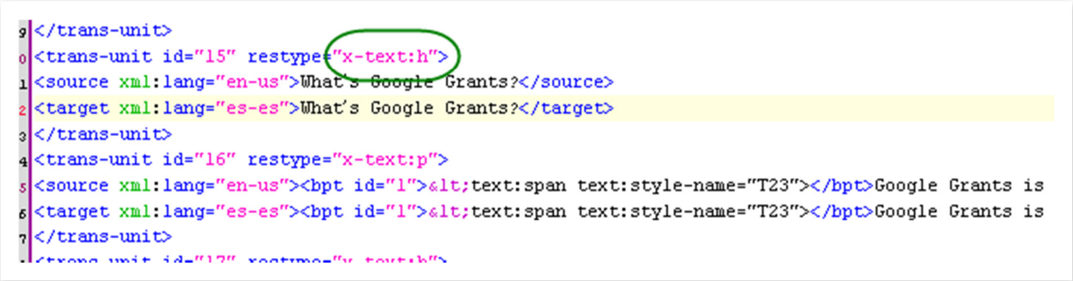


Code	Description
CO	Text embedded in an image.
FLD	Document field or placeholder text.
FN	Footnote text.
H	Heading text.
KW	Keyword list entry, such as an index entry, for example.
LI	Item from a bulleted or numbered list.
MP	Master page text.
PF	Page footer text.
PH	Page header text.
P	Paragraph text.
REF	Reference to a related paragraph.
S	Script. This is translatable text inside a piece of code.
SB	Sidebar text.
TD	Table cell text.
TH	Table heading text.
T	Translatable tag content.

Figure 41 - Structure codes for SDL Trados Studio

4.2.3.2. Text Editor

The XLIFF implementation of jEdit shows information about headings by adding a <restype="x-text:h"> attribute to the trans-unit element. The attribute "restype" (resource type) indicates the resource type of the element.



```
</trans-unit>
<trans-unit id="15" restype="x-text:h">
<source xml:lang="en-us">What's Google Grants?</source>
<target xml:lang="es-es">What's Google Grants?</target>
</trans-unit>
<trans-unit id="16" restype="x-text:p">
<source xml:lang="en-us"><bpt id="1">&lt;text:span text:style-name="T23"></bpt>Google Grants is
<target xml:lang="es-es"><bpt id="1">&lt;text:span text:style-name="T23"></bpt>Google Grants is
</trans-unit>
<trans-unit id="17" restype="x-text:h">
```

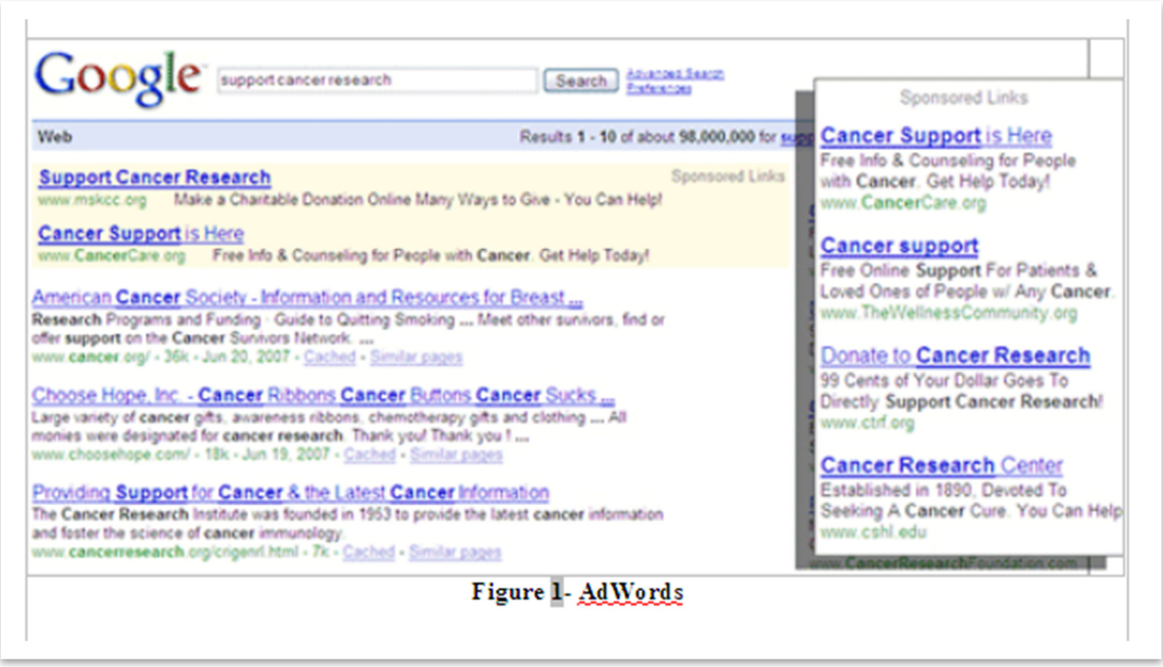
Figure 42 - Headings in jEdit

4.2.4. Images and captions[10]

4.2.4.1. XLIFF editors

Images are valuable sources of information for the translator as they typically reflect what the text is describing and therefore they can help them understand the text better. They often contain inserted text, in which case it would be necessary to localise the text using an image localisation tool.

Figure 43 shows an image and its caption from a source file in OpenDocument format.



Results 1 - 10 of about 98,000,000 for supp

Sponsored Links

Support Cancer Research
www.mskcc.org Make a Charitable Donation Online Many Ways to Give - You Can Help!

Cancer Support is Here
www.CancerCare.org Free Info & Counseling for People with Cancer. Get Help Today!

American Cancer Society - Information and Resources for Breast ...
Research Programs and Funding · Guide to Quitting Smoking ... Meet other survivors, find or offer support on the Cancer Survivors Network. ...
www.cancer.org/ - 36k - Jun 20, 2007 - Cached - Similar pages

Choose Hope, Inc. - Cancer Ribbons Cancer Buttons Cancer Sucks ...
Large variety of cancer gifts, awareness ribbons, chemotherapy gifts and clothing ... All monies were designated for cancer research. Thank you! Thank you ! ...
www.choosehope.com/ - 18k - Jun 19, 2007 - Cached - Similar pages

Providing Support for Cancer & the Latest Cancer Information
The Cancer Research Institute was founded in 1953 to provide the latest cancer information and foster the science of cancer immunology.
www.cancerresearch.org/crigenrl.html - 7k - Cached - Similar pages

Sponsored Links

Cancer Support is Here
Free Info & Counseling for People with Cancer. Get Help Today!
www.CancerCare.org

Cancer support
Free Online Support For Patients & Loved Ones of People w/ Any Cancer.
www.TheWellnessCommunity.org

Donate to Cancer Research
99 Cents of Your Dollar Goes To Directly Support Cancer Research!
www.ctrf.org

Cancer Research Center
Established in 1890, Devoted To Seeking A Cancer Cure. You Can Help
www.cshl.edu

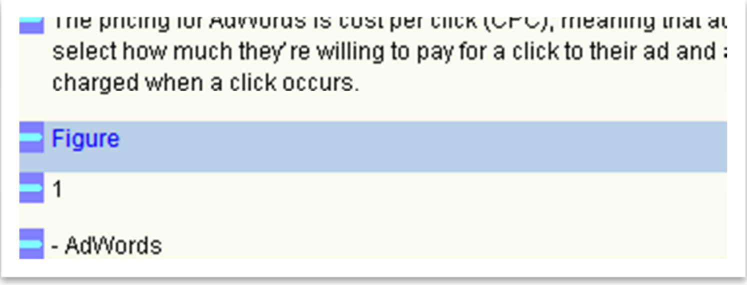
www.CancerResearchFoundation.com

Figure 1- AdWords

Figure 43 - Image with caption in OpenDocument

The XLIFF implementation of Open Language Tools does not provide any reference to the image from figure 43, although the caption is included in the text to be translated. As a consequence, in the case of figure 44, translators would not know that there is an image in the document unless they have access to the source file. They would only be able to see the caption of the image.

[10] A caption is a few lines of text used to explain or elaborate on photographs.



The pricing for AdWords is cost per click (CPC), meaning that ad
select how much they're willing to pay for a click to their ad and
charged when a click occurs.

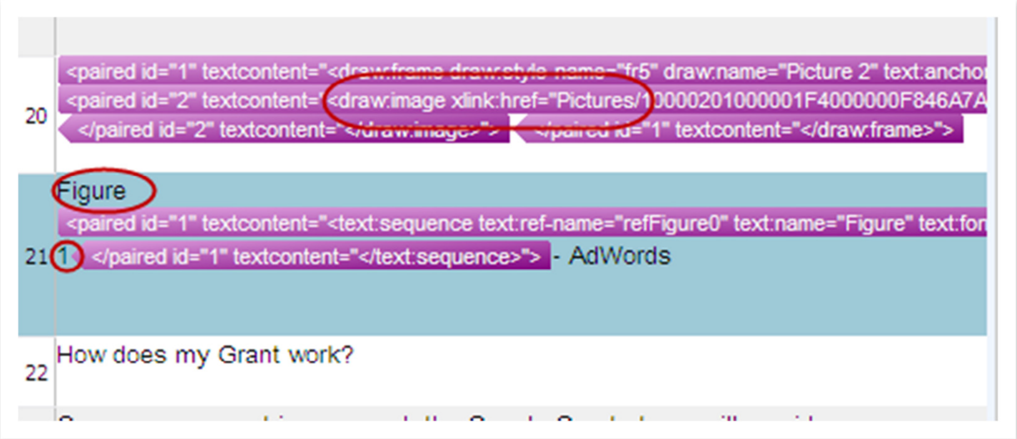
Figure

1

- AdWords

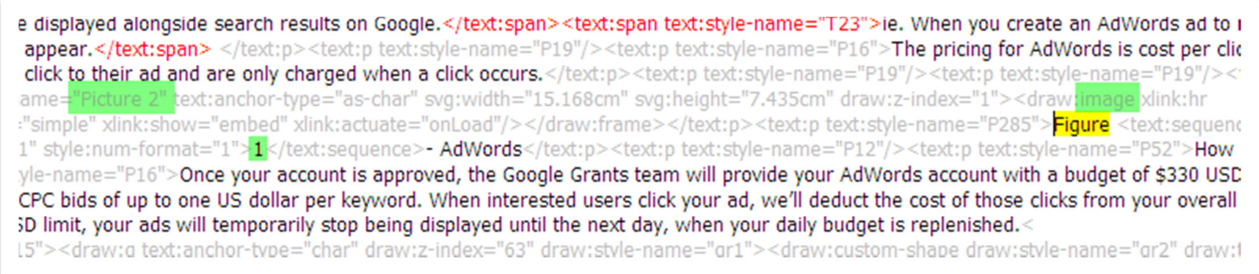
Figure 44 - Images/Captions in Open Language Tools

SDL Trados and Translation Workspace display some elements, which inform the translator about the embedded pictures in the document, although the images themselves cannot be seen. See below examples of how the source image from figure 43 is displayed by both tools.



```
20  <paired id="1" textcontent="<draw:frame draw:style-name="fr5" draw:name="Picture 2" text:ancho
    <paired id="2" textcontent="<draw:image xlink:href="Pictures/10000201000001F4000000F846A7A
    </paired id="2" textcontent="</draw:image>">  </paired id="1" textcontent="</draw:frame>">

    Figure
    <paired id="1" textcontent="<text:sequence text:ref-name="refFigure0" text:name="Figure" text:for
21  1 </paired id="1" textcontent="</text:sequence>"> - AdWords

    How does my Grant work?
22
```

Figure 45 - Images in SDL Trados Studio

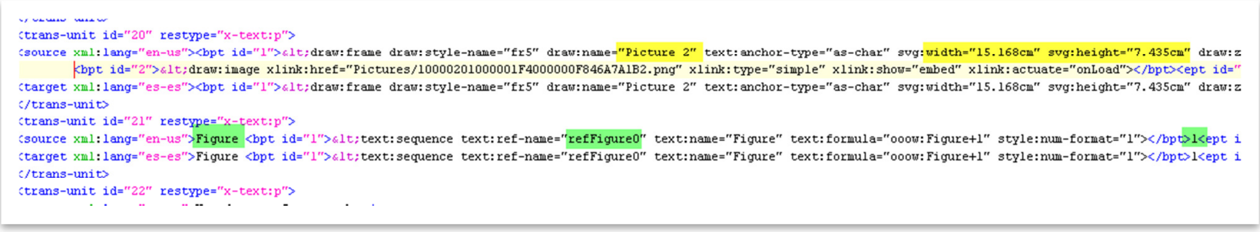


```
e displayed alongside search results on Google.</text:span><text:span text:style-name="T23">ie. When you create an AdWords ad to
 appear.</text:span> </text:p><text:p text:style-name="P19"/><text:p text:style-name="P16">The pricing for AdWords is cost per clic
 click to their ad and are only charged when a click occurs.</text:p><text:p text:style-name="P19"/><text:p text:style-name="P19"/><
ame="Picture 2" text:anchor-type="as-char" svg:width="15.168cm" svg:height="7.435cm" draw:z-index="1"><draw:image xlink:hr
:"simple" xlink:show="embed" xlink:actuate="onLoad"/></draw:frame></text:p><text:p text:style-name="P285">Figure <text:sequen
1" style:num-format="1">1</text:sequence>- AdWords</text:p><text:p text:style-name="P12"/><text:p text:style-name="P52">How
yle-name="P16">Once your account is approved, the Google Grants team will provide your AdWords account with a budget of $330 USD
CPC bids of up to one US dollar per keyword. When interested users click your ad, we'll deduct the cost of those clicks from your overall
SD limit, your ads will temporarily stop being displayed until the next day, when your daily budget is replenished.<
15"><draw:g text:anchor-type="char" draw:z-index="63" draw:style-name="gr1"><draw:custom-shape draw:style-name="gr2" draw:
```

Figure 46 - Images in Translation Workspace

4.2.4.2 Text editor

When opening the XLIFF file corresponding to the source image from figure 43, the result is the following:



```
<trans-unit id="20" restype="x-text:p">
<source xml:lang="en-us"><bpt id="1">&lt;draw:frame draw:style-name="fr5" draw:name="Picture 2" text:anchor-type="as-char" svg:width="15.168cm" svg:height="7.435cm" draw:z
        <bpt id="2">&lt;draw:image xlink:href="Pictures/10000201000001F4000000F846A7A1B2.png" xlink:type="simple" xlink:show="embed" xlink:actuate="onLoad"></bpt><ept id="
<target xml:lang="es-es"><bpt id="1">&lt;draw:frame draw:style-name="fr5" draw:name="Picture 2" text:anchor-type="as-char" svg:width="15.168cm" svg:height="7.435cm" draw:z
</trans-unit>
<trans-unit id="21" restype="x-text:p">
<source xml:lang="en-us">Figure <bpt id="1">&lt;text:sequence text:ref-name="refFigure0" text:name="Figure" text:formula="ooow:Figure+1" style:num-format="1"></bpt>1<ept i
<target xml:lang="es-es">Figure <bpt id="1">&lt;text:sequence text:ref-name="refFigure0" text:name="Figure" text:formula="ooow:Figure+1" style:num-format="1"></bpt>1<ept i
</trans-unit>
<trans-unit id="22" restype="x-text:p">
```

Figure 47 - images in jEdit

Therefore, XLIFF indicates that there is an image and provides details about it, such as the width and height (see yellow highlight). XLIFF also specifies information about the caption with the attribute “refFigure0” (see green highlight).

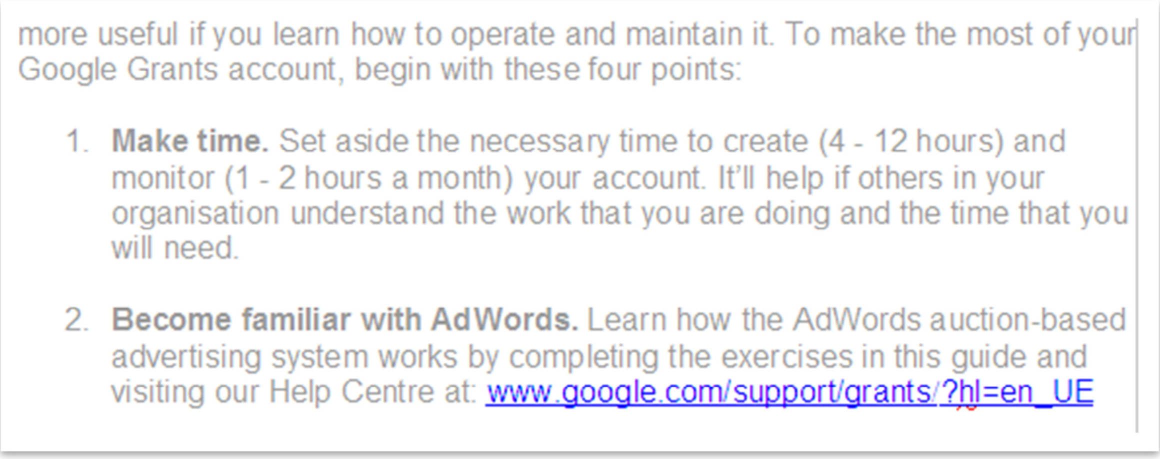
Ideally, the translator would be able to see the image, not just a reference to it. This is possible if the translator has access to the source file.

4.2.5. Lists

4.2.5.1. XLIFF editors

Numbered lists are not flagged with Open Language Tools or with SDL Trados Studio, and as a result, translators cannot possibly know that they are translating a numbered list of elements without seeing the source file.

Figure 48 is an example of a source file (OpenDocument), which includes a numbered list of elements.



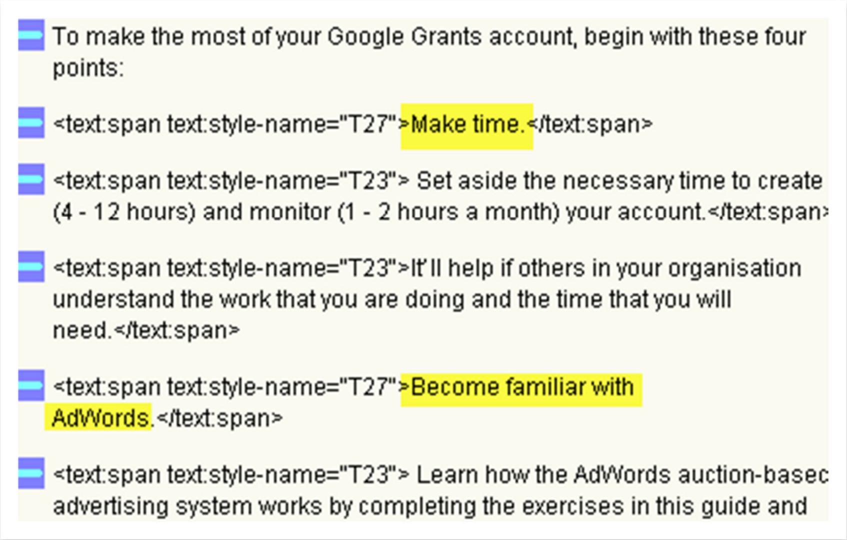
more useful if you learn how to operate and maintain it. To make the most of your Google Grants account, begin with these four points:

1. **Make time.** Set aside the necessary time to create (4 - 12 hours) and monitor (1 - 2 hours a month) your account. It'll help if others in your organisation understand the work that you are doing and the time that you will need.

2. **Become familiar with AdWords.** Learn how the AdWords auction-based advertising system works by completing the exercises in this guide and visiting our Help Centre at: www.google.com/support/grants/?hl=en_UE

Figure 48 - Lists in OpenDocument

The XLIFF implementation of Open Language Tools and SDL Trados Studio do not incorporate any information specifying the numbered list of elements from the source text. Instead they only indicate that the text has a certain style (T23, T27) through the <span> tag. A span tag provides no visual change by itself, but enables the linking of the text included in the span element to a certain text style. However, it would be difficult for the translator to know what each style code stands for as they would have to go back to the skeleton of the XLIFF file and look up what T23 and T27 correspond to. Each XLIFF file creates their style codes to reference a particular style.



To make the most of your Google Grants account, begin with these four points:

<text:span text:style-name="T27">Make time.</text:span>

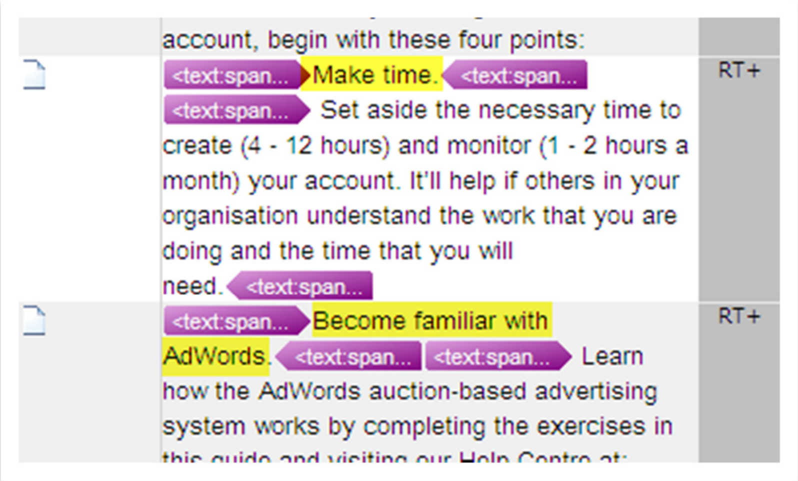
<text:span text:style-name="T23"> Set aside the necessary time to create (4 - 12 hours) and monitor (1 - 2 hours a month) your account.</text:span>

<text:span text:style-name="T23">It'll help if others in your organisation understand the work that you are doing and the time that you will need.</text:span>

<text:span text:style-name="T27">Become familiar with AdWords.</text:span>

<text:span text:style-name="T23"> Learn how the AdWords auction-basec advertising system works by completing the exercises in this guide and

Figure 49 - Lists in Open Language Tools



account, begin with these four points:

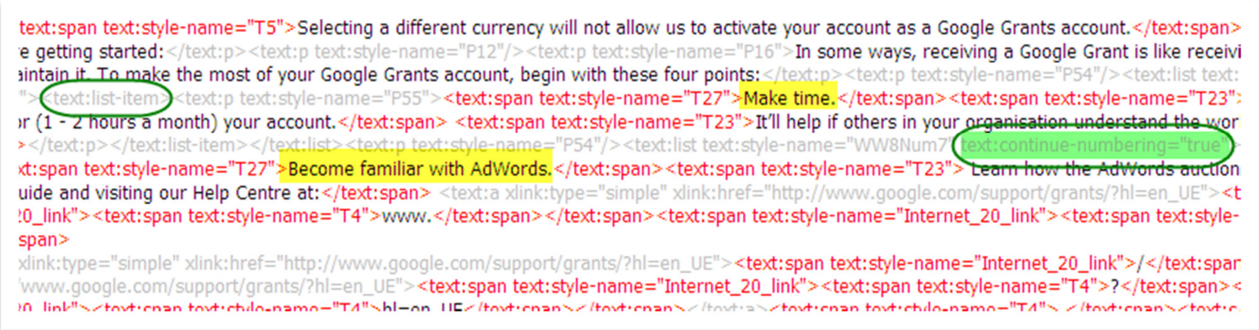
<text:span... Make time. <text:span... RT+

<text:span... Set aside the necessary time to create (4 - 12 hours) and monitor (1 - 2 hours a month) your account. It'll help if others in your organisation understand the work that you are doing and the time that you will need. <text:span...

<text:span... Become familiar with AdWords. <text:span... <text:span... Learn how the AdWords auction-based advertising system works by completing the exercises in this guide and visiting our Help Centre at: RT+

Figure 50 - Lists in SDL Trados Studio

Translation Workspace flags the list of elements with the tag <text:list-item> although it does not specify the numbers of the lists (see green circle in figure 51). Translators would know that there is a numbered list (see green circle with the green highlighted text from figure 51), but it is very difficult to distinguish when the list starts or ends.

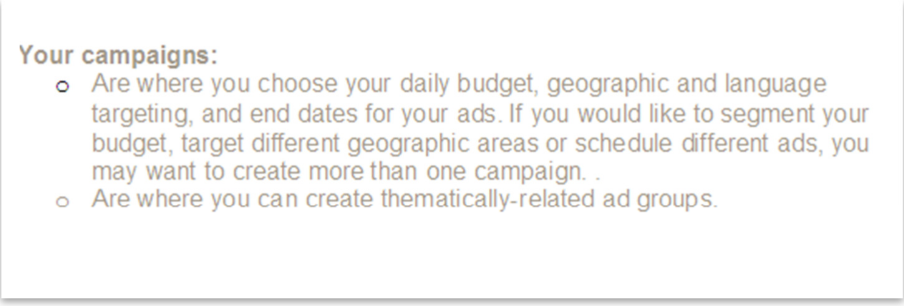


```
text:span text:style-name="T5">Selecting a different currency will not allow us to activate your account as a Google Grants account.</text:span>
e getting started:</text:p><text:p text:style-name="P12"/><text:p text:style-name="P16">In some ways, receiving a Google Grant is like receivi
aintain it. To make the most of your Google Grants account, begin with these four points:</text:p><text:p text:style-name="P54"/><text:list text:
"><text:list-item><text:p text:style-name="P55"><text:span text:style-name="T27">Make time.</text:span><text:span text:style-name="T23":
or (1 - 2 hours a month) your account.</text:span> <text:span text:style-name="T23">It'll help if others in your organisation understand the wor
></text:p></text:list-item></text:list><text:p text:style-name="P54"/><text:list text:style-name="WW8Num7" text:continue-numbering="true">
xt:span text:style-name="T27">Become familiar with AdWords.</text:span><text:span text:style-name="T23"> Learn how the AdWords auction
uide and visiting our Help Centre at:</text:span> <text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants/?hl=en_UE"><t
20_link"><text:span text:style-name="T4">www.</text:span></text:span><text:span text:style-name="Internet_20_link"><text:span text:style-
span>
xlink:type="simple" xlink:href="http://www.google.com/support/grants/?hl=en_UE"><text:span text:style-name="Internet_20_link">/</text:span
'www.google.com/support/grants/?hl=en_UE"><text:span text:style-name="Internet_20_link"><text:span text:style-name="T4">?</text:span><
20_link"><text:span text:style-name="T4">hl=en_UE</text:span></text:span></text:a><text:span text:style-name="T4">.</text:span><text:s
```

Figure 51 - Lists in Translation Workspace

Bullet point lists are displayed neither with Open Language Tools nor with SDL Trados Studio. However, they are captured in Translation Workspace using the <text:list-item> as the numbered list. The only difference is that there is not an attribute specifying the numbering (see the green highlighted circle from figure 51).

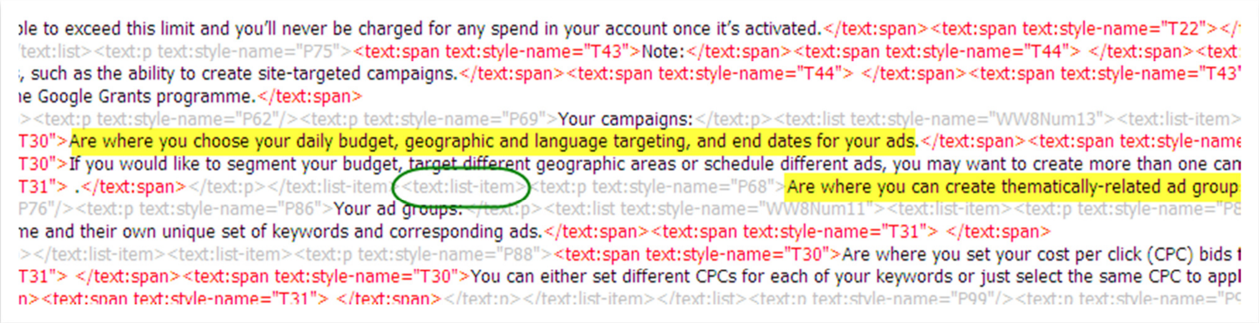
Figure 52 shows a bullet point list of elements in a source text and figure 53 shows the same list but converted into XLIFF and opened with Translation Workspace. The <text:list-item> attribute can be seen.



Your campaigns:

- Are where you choose your daily budget, geographic and language targeting, and end dates for your ads. If you would like to segment your budget, target different geographic areas or schedule different ads, you may want to create more than one campaign. .
- Are where you can create thematically-related ad groups.

Figure 52 - Bullet point list in OpenDocument

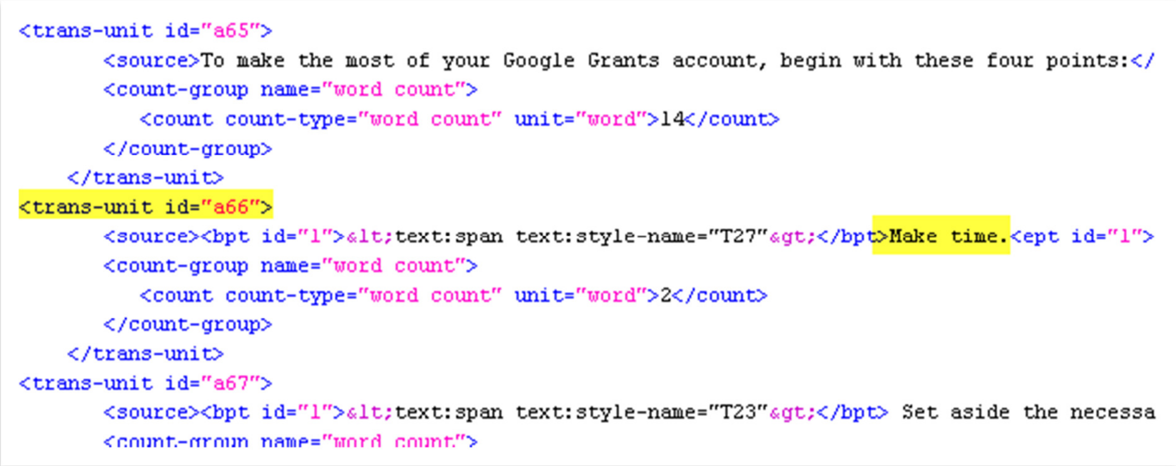


```
ɔle to exceed this limit and you'll never be charged for any spend in your account once it's activated.</text:span><text:span text:style-name="T22"></
'text:list><text:p text:style-name="P75"><text:span text:style-name="T43">Note:</text:span><text:span text:style-name="T44"> </text:span><text:
, such as the ability to create site-targeted campaigns.</text:span><text:span text:style-name="T44"> </text:span><text:span text:style-name="T43"
ıe Google Grants programme.</text:span>
></text:p text:style-name="P62"/><text:p text:style-name="P69">Your campaigns:</text:p><text:list text:style-name="WW8Num13"><text:list-item>
T30">Are where you choose your daily budget, geographic and language targeting, and end dates for your ads.</text:span><text:span text:style-name
T30">If you would like to segment your budget, target different geographic areas or schedule different ads, you may want to create more than one can
T31"> .</text:span></text:p></text:list-item><text:list-item><text:p text:style-name="P68">Are where you can create thematically-related ad group
P76"/><text:p text:style-name="P86">Your ad groups:</text:p><text:list text:style-name="WW8Num11"><text:list-item><text:p text:style-name="P8
ne and their own unique set of keywords and corresponding ads.</text:span><text:span text:style-name="T31"> </text:span>
></text:list-item><text:list-item><text:p text:style-name="P88"><text:span text:style-name="T30">Are where you set your cost per click (CPC) bids f
T31"> </text:span><text:span text:style-name="T30">You can either set different CPCs for each of your keywords or just select the same CPC to appl
n><text:span text:style-name="T31"> </text:span></text:p></text:list-item></text:list><text:p text:style-name="P99"/><text:p text:style-name="P9
```

Figure 53 - Bullet points in Translation Worskpace

4.2.5.2. Text editors

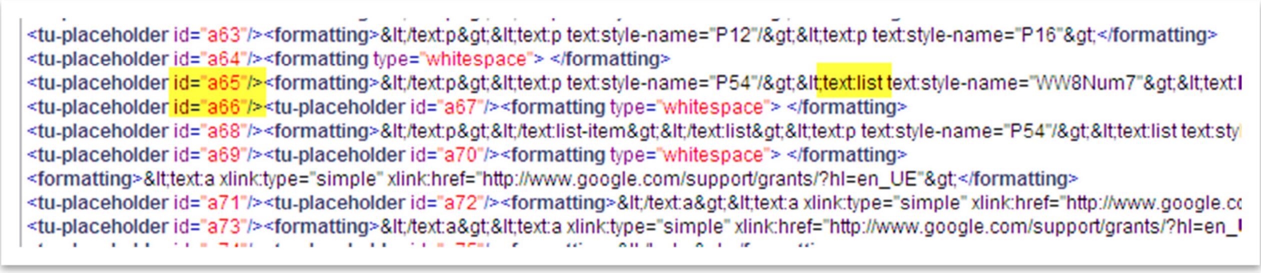
The XLIFF file opened with jEdit does not show any indications about the numbered list from the source file (figure 48), as it can be seen below in figure 54 below. Note that the translation unit of the first element of the list is "a66".



```
<trans-unit id="a65">
       <source>To make the most of your Google Grants account, begin with these four points:</
       <count-group name="word count">
          <count count-type="word count" unit="word">14</count>
       </count-group>
    </trans-unit>
<trans-unit id="a66">
       <source><bpt id="1">&lt;text:span text:style-name="T27"&gt;</bpt>Make time.<ept id="1">
       <count-group name="word count">
          <count count-type="word count" unit="word">2</count>
       </count-group>
    </trans-unit>
<trans-unit id="a67">
       <source><bpt id="1">&lt;text:span text:style-name="T23"&gt;</bpt> Set aside the necessa
       <count-group name="word count">
```

Figure 54 - Lists in jEdit

When looking at the corresponding skeleton of the XLIFF file it can be seen that the translation unit "a65" includes a "text:list" attribute, which indicates that there is a list of elements. It is interesting that the "text:list" attribute is placed in the translation unit before the first element of the list ("a66"). This seems to be the case in all lists of elements evaluated in this Case Study.



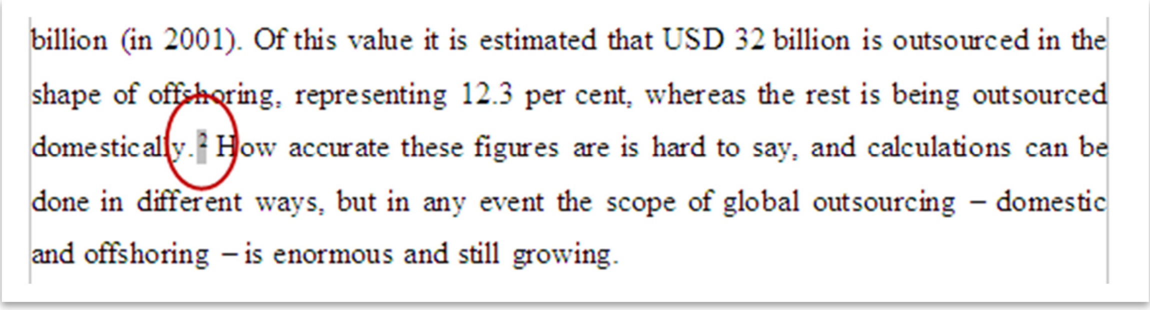
```
<tu-placeholder id="a63"/><formatting>&lt;/text:p&gt;&lt;text:p text:style-name="P12"/&gt;&lt;text:p text:style-name="P16"&gt;</formatting>
<tu-placeholder id="a64"/><formatting type="whitespace"> </formatting>
<tu-placeholder id="a65"/><formatting>&lt;/text:p&gt;&lt;text:p text:style-name="P54"/&gt;&lt;text:list text:style-name="WW8Num7"&gt;&lt;text:l
<tu-placeholder id="a66"/><tu-placeholder id="a67"/><formatting type="whitespace"> </formatting>
<tu-placeholder id="a68"/><formatting>&lt;/text:p&gt;&lt;/text:list-item&gt;&lt;/text:list&gt;&lt;text:p text:style-name="P54"/&gt;&lt;text:list text:sty
<tu-placeholder id="a69"/><tu-placeholder id="a70"/><formatting type="whitespace"> </formatting>
<formatting>&lt;text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants/?hl=en_UE"&gt;</formatting>
<tu-placeholder id="a71"/><tu-placeholder id="a72"/><formatting>&lt;/text:a&gt;&lt;text:a xlink:type="simple" xlink:href="http://www.google.co
<tu-placeholder id="a73"/><formatting>&lt;/text:a&gt;&lt;text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants/?hl=en_
```

Figure 55 - Lists in skeleton

4.2.6. Footnotes

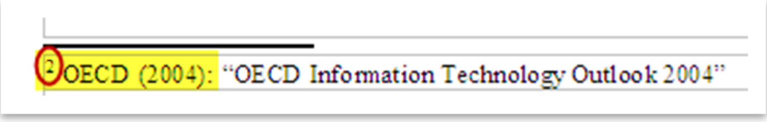
4.2.6.1. XLIFF editors

To see how footnotes are displayed by XLIFF it is necessary to give an example of a footnote in a source text (figures 56 and 57).



billion (in 2001). Of this value it is estimated that USD 32 billion is outsourced in the shape of offshoring, representing 12.3 per cent, whereas the rest is being outsourced domestically.[2] How accurate these figures are is hard to say, and calculations can be done in different ways, but in any event the scope of global outsourcing – domestic and offshoring – is enormous and still growing.

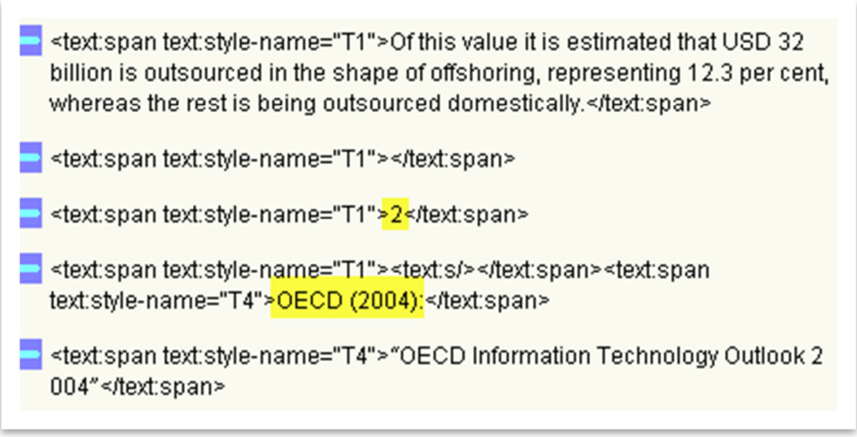
Figure 56 - Footnotes in OpenDocument



[2]OECD (2004): "OECD Information Technology Outlook 2004"

Figure 57 - Footnotes in OpenDocument 2

Footnotes are not labeled with Open Language Tools. They are displayed embedded in the text, just after their corresponding footnote number. Translators facing the translation of this text would probably feel very confused if they do not have access to the source file.

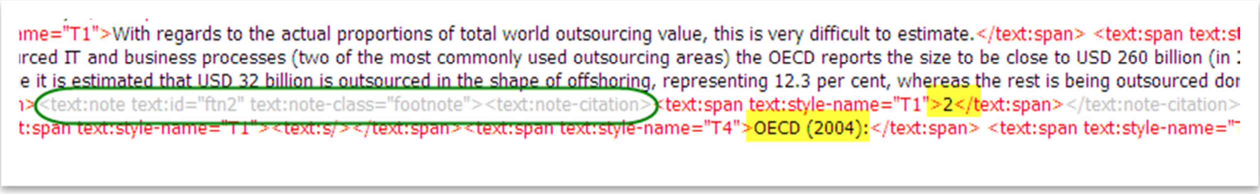


```
<text:span text:style-name="T1">Of this value it is estimated that USD 32 billion is outsourced in the shape of offshoring, representing 12.3 per cent, whereas the rest is being outsourced domestically.</text:span>

<text:span text:style-name="T1"></text:span>

<text:span text:style-name="T1">2</text:span>

<text:span text:style-name="T1"><text:s/></text:span><text:span text:style-name="T4">OECD (2004):</text:span>

<text:span text:style-name="T4">"OECD Information Technology Outlook 2004"</text:span>
```

Figure 58 - Footnotes in Open Language Tools

The XLIFF implementation used by Translation Workspace and SDL Trados Studio indicates footnotes by adding an attribute to the text. These two programs also embed the footnotes in the text, just after or before the footnote number.

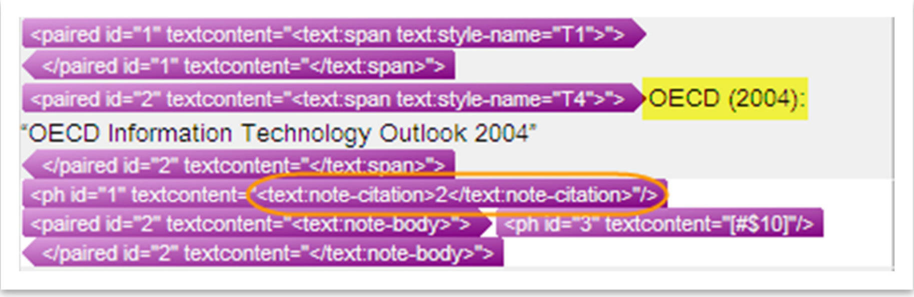
Translation Workspace marks the footnote with the value "fn2", the attribute "footnote" and the element <text:note-citation>, which can be seen in the green circle in figure 59.



```
ıme="T1">With regards to the actual proportions of total world outsourcing value, this is very difficult to estimate.</text:span> <text:span text:st
ırced IT and business processes (two of the most commonly used outsourcing areas) the OECD reports the size to be close to USD 260 billion (in :
e it is estimated that USD 32 billion is outsourced in the shape of offshoring, representing 12.3 per cent, whereas the rest is being outsourced dor
ı><text:note text:id="ftn2" text:note-class="footnote"><text:note-citation><text:span text:style-name="T1">2</text:span></text:note-citation>
t:span text:style-name="T1"><text:s/></text:span><text:span text:style-name="T4">OECD (2004):</text:span> <text:span text:style-name="
```

Figure 59 - Footnotes in Translation Workspace

The XLIFF implementation of SDL Trados Studio incorporates a <text:note-citation> element to the footnote number after the footnote itself, as can be seen in Figure 60. The footnote number is inserted in the <text:note-citation>, which means that the translator needs to expand the tag view in order to see the footnote number. Note that tags can be partial and full (the screenshot shows full tag text).



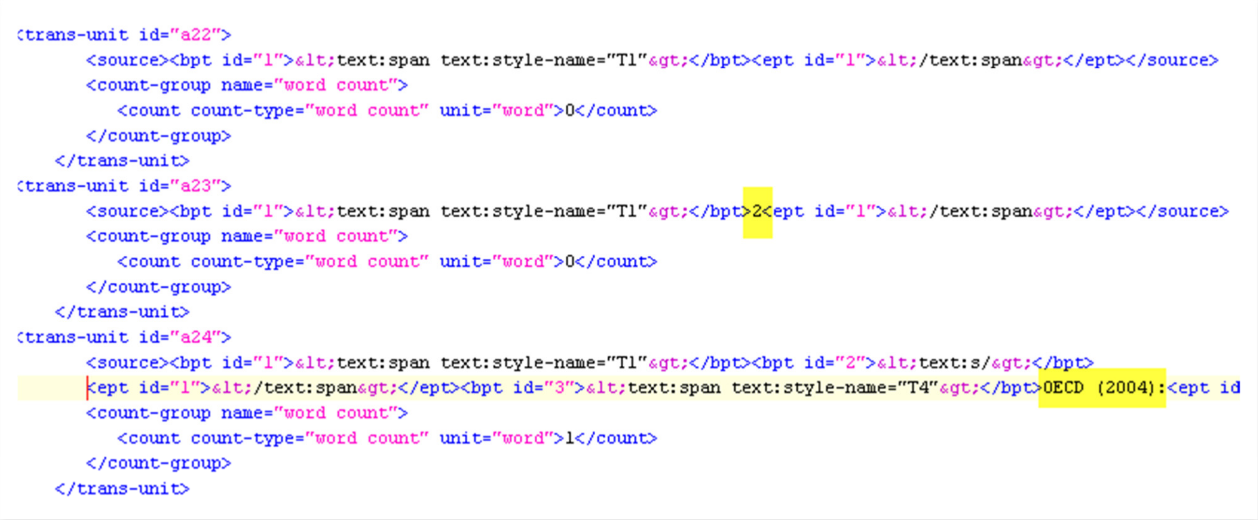
```
<paired id="1" textcontent="<text:span text:style-name="T1">">
</paired id="1" textcontent="</text:span>">
<paired id="2" textcontent="<text:span text:style-name="T4">">OECD (2004):
"OECD Information Technology Outlook 2004"
</paired id="2" textcontent="</text:span>">
<ph id="1" textcontent="<text:note-citation>2</text:note-citation>"/>
<paired id="2" textcontent="<text:note-body>"> <ph id="3" textcontent="[#$10]"/>
</paired id="2" textcontent="</text:note-body>">
```

Figure 60 - Footnote in SDL Trados Studio

Having footnotes inserted in the text is not recommended as it can create serious problems for translators. If they do not have access to the source text and are unable to understand XLIFF tags it is very likely that they will not know what they are translating.

4.2.6.2. Text editors

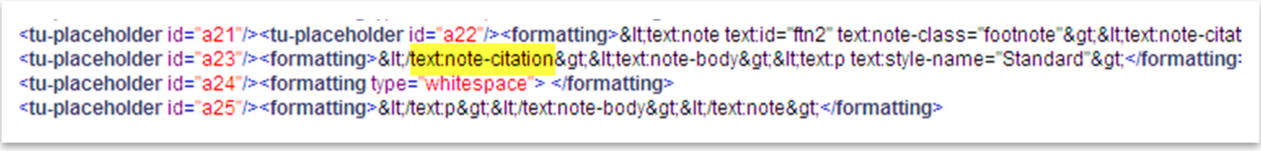
When observing the XLIFF file with jEdit one can see that the footnote is embedded within the text after the footnote number, just as in the previous cases. Note that the footnote number is the translation unit "a23".



```
<trans-unit id="a22">
        <source><bpt id="1">&lt;text:span text:style-name="T1"&gt;</bpt><ept id="1">&lt;/text:span&gt;</ept></source>
        <count-group name="word count">
           <count count-type="word count" unit="word">0</count>
        </count-group>
    </trans-unit>
<trans-unit id="a23">
        <source><bpt id="1">&lt;text:span text:style-name="T1"&gt;</bpt>2<ept id="1">&lt;/text:span&gt;</ept></source>
        <count-group name="word count">
           <count count-type="word count" unit="word">0</count>
        </count-group>
    </trans-unit>
<trans-unit id="a24">
        <source><bpt id="1">&lt;text:span text:style-name="T1"&gt;</bpt><bpt id="2">&lt;text:s/&gt;</bpt>
        <ept id="1">&lt;/text:span&gt;</ept><bpt id="3">&lt;text:span text:style-name="T4"&gt;</bpt>OECD (2004):<ept id
        <count-group name="word count">
           <count count-type="word count" unit="word">1</count>
        </count-group>
    </trans-unit>
```

Figure 61 - Footnotes in jEdit

The XLIFF file shows no tags indicating that the inserted element is a footnote. However, if the skeleton file is opened one can see that the trans-unit "a23" makes a reference to a footnote. Therefore, XLIFF does have the ability to show footnotes.



```
<tu-placeholder id="a21"/><tu-placeholder id="a22"/><formatting>&lt;text:note text:id="ftn2" text:note-class="footnote"&gt;&lt;text:note-citat
<tu-placeholder id="a23"/><formatting>&lt;/text:note-citation&gt;&lt;text:note-body&gt;&lt;text:p text:style-name="Standard"&gt;</formatting:
<tu-placeholder id="a24"/><formatting type="whitespace"> </formatting>
<tu-placeholder id="a25"/><formatting>&lt;/text:p&gt;&lt;/text:note-body&gt;&lt;/text:note&gt;</formatting>
```

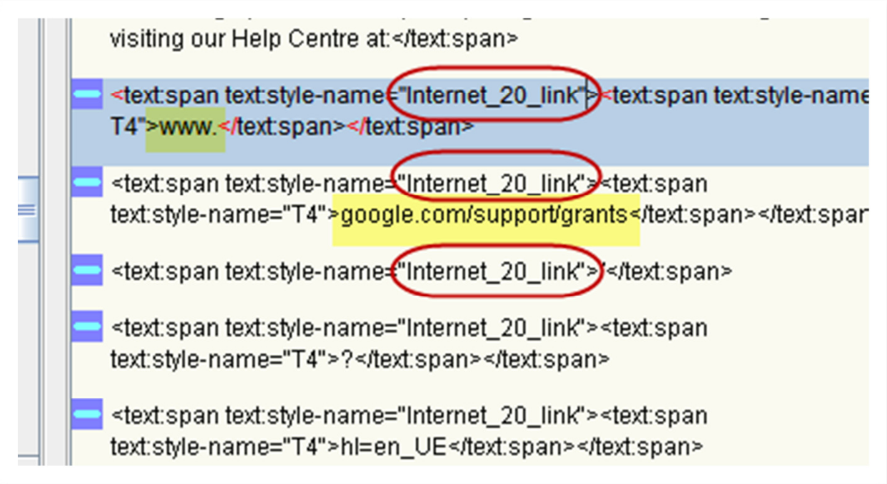
Figure 62 - Footnote in Skeleton

4.2.7. Links

4.2.7.1. XLIFF editors

Links are flagged by the three XLIFF editors used in this Case Study.

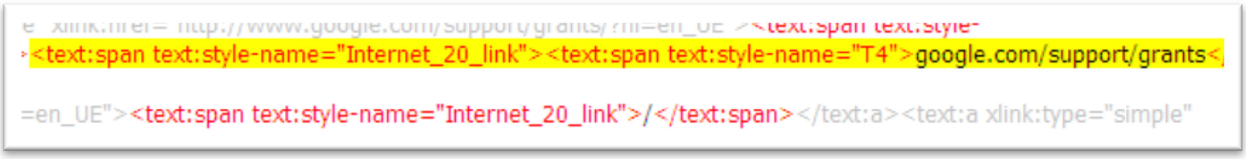
The XLIFF implementation of Open Language Tools modifies the <text> element with the attribute "Internet_20_link".



```
visiting our Help Centre at:</text:span>

<text:span text:style-name="Internet_20_link"><text:span text:style-name
T4">www.</text:span></text:span>

<text:span text:style-name="Internet_20_link"><text:span
text:style-name="T4">google.com/support/grants</text:span></text:spar

<text:span text:style-name="Internet_20_link">'</text:span>

<text:span text:style-name="Internet_20_link"><text:span
text:style-name="T4">?</text:span></text:span>

<text:span text:style-name="Internet_20_link"><text:span
text:style-name="T4">hl=en_UE</text:span></text:span>
```

Figure 63 - Links in Open Language Tools

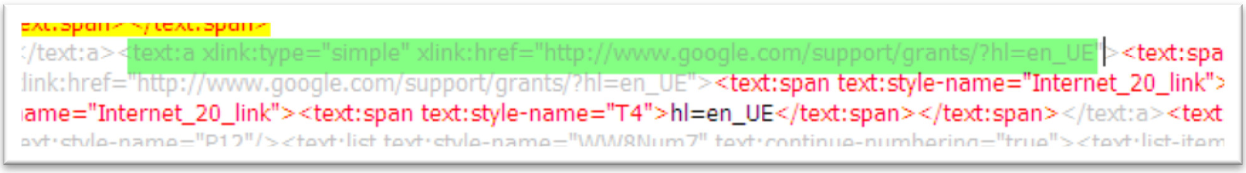
Translation Workspace adds a similar attribute "Internet_20_link" to the <text> element:



```
e" xlink:href="http://www.google.com/support/grants/?hl=en_UE"><text:span text:style-
><text:span text:style-name="Internet_20_link"><text:span text:style-name="T4">google.com/support/grants<

=en_UE"><text:span text:style-name="Internet_20_link">/</text:span></text:a><text:a xlink:type="simple"
```

Figure 64 - Links1 in Translation Workspace

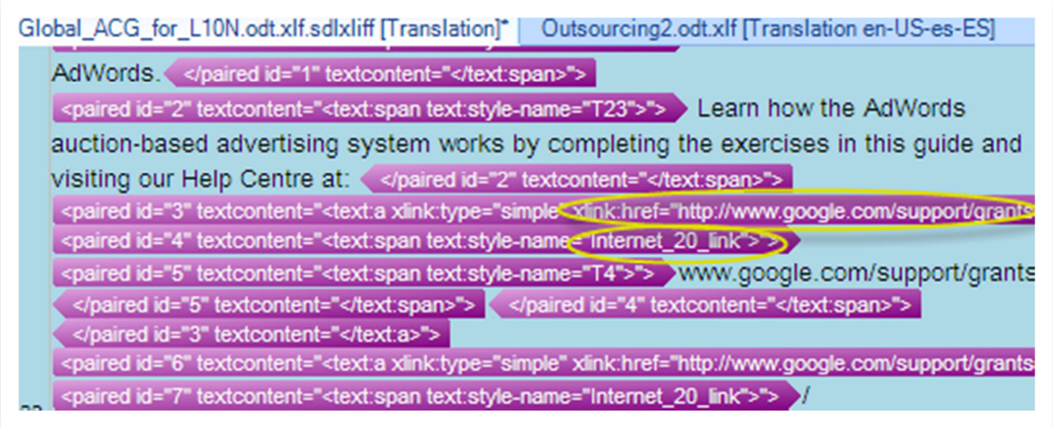
In addition, it incorporates a <xlink> element where the source URL can be seen in a non-editable form, which means that the tag is protected.



```
ext:span></text:span>
</text:a><text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants/?hl=en_UE"><text:spa
link:href="http://www.google.com/support/grants/?hl=en_UE"><text:span text:style-name="Internet_20_link">
ame="Internet_20_link"><text:span text:style-name="T4">hl=en_UE</text:span></text:span></text:a><text
ext:style-name="P12"/><text:list text:style-name="WW8Num7" text:continue-numbering="true"><text:list-item
```

Figure 65 - Links2 in Translation Workspace

SDL Trados Studio also adds the attribute “Internet_20_link” to the <text> element and the <xlink> element in a protected tag, as can be seen in the figure 63 below.



Global_ACG_for_L10N.odt.xlf.sdlxliff [Translation]*
Outsourcing2.odt.xlf [Translation en-US-es-ES]

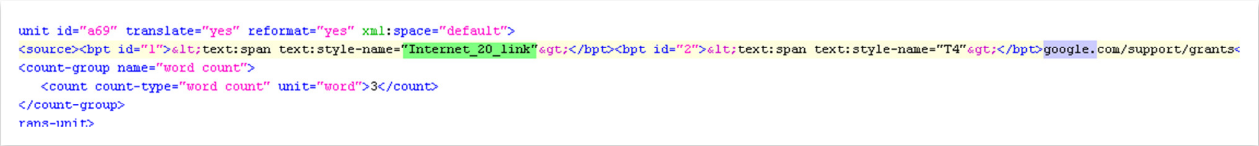
```
AdWords. </paired id="1" textcontent="</text:span>">
<paired id="2" textcontent="<text:span text:style-name="T23">"> Learn how the AdWords
auction-based advertising system works by completing the exercises in this guide and
visiting our Help Centre at: </paired id="2" textcontent="</text:span>">
<paired id="3" textcontent="<text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants
<paired id="4" textcontent="<text:span text:style-name="Internet_20_link">">
<paired id="5" textcontent="<text:span text:style-name="T4">"> www.google.com/support/grants
</paired id="5" textcontent="</text:span>"> </paired id="4" textcontent="</text:span>">
</paired id="3" textcontent="</text:a>">
<paired id="6" textcontent="<text:a xlink:type="simple" xlink:href="http://www.google.com/support/grants
<paired id="7" textcontent="<text:span text:style-name="Internet_20_link">"> /
```

Figure 66 - Links in SDL Trados Studio

Therefore, XLIFF editors seem to be quite consistent in the way they display links or URLs, which are fairly easy to identify in most cases.

4.2.7.2. Text editors

The XLIFF implementation of the text editor jEdit also uses the attribute "Internet_20_link" to the <text> element to indicate URLs.



```
unit id="a69" translate="yes" reformat="yes" xml:space="default">
<source><bpt id="1">&lt;text:span text:style-name="Internet_20_link"&gt;</bpt><bpt id="2">&lt;text:span text:style-name="T4"&gt;</bpt>google.com/support/grants<
<count-group name="word count">
   <count count-type="word count" unit="word">3</count>
</count-group>
rans-unit>
```

Figure 67 - Links in jEdit

4.3. Questionnaire

The questionnaire has been distributed in order to evaluate the usage of XLIFF and to find out about the most common missing information in translation projects.

The response volume has been twenty-seven replies, out of a much higher number. The questionnaire was sent to translators of the distribution lists and associations stated in the "Methodology" section, although it is difficult to confirm how many people actually received the questionnaire. As an estimate, more than 500 people received an email containing the questionnaire. Nevertheless, twenty-seven replies is a relevant number for the purposes of this research project.

Unfortunately, some of the answers provided were not valid. In some occasions, translators did not fully understand the questions and therefore their answers were not proving; in other occasions, subjects were unable to answer some of the questions.

As mentioned earlier, the questionnaire was composed of five questions, which are exposed below with a summary of the replies. Not all answers are captured in the following sections; only conclusive answers in terms of number of repetitions and relevance to the topic. Full answers can be consulted in Appendix B[11].

1. What programs do you use for the translation of documentation/online help files?

[11] Appendixes have been removed from this version for publication due to confidentiality issues.

Twenty-five out the twenty-seven people use SDL Trados to translate documentation files. This is not surprising given the fact that SDL Trados is a leading translation tool and that the last release has incorporated filters to treat new formats such as XLIFF. Some of the subjects answered that, in particular, they use TagEditor, SDLX or Translator Workbench, which are also SDL Trados products.

The second most used tool is Translation Workspace, previously called Logoport, which is a Lionbridge propriety tool. This fact is not unexpected, because Lionbridge is one of the largest translation companies and hire a large number of freelance translators who are required to use their tool. Eight people replied that they use Translation Workspace.

Microsoft Word and Excel are used by five and four people respectively, which in the past would probably have been a higher number of users. Nowadays, translators use translation memory tools, and therefore, if they get a Word file to translate, it will be processed by a translation program, such as SDL Trados.

Three people answered that they use XLIFF as a program, which is a wrong answer, because XLIFF is not a program but a format. They probably meant that they use XLIFF editors.

2. What kind of information do these programs allow you to see?

Thirteen out of twenty-seven subjects replied that they are allowed to see the format of the document, and seven people answered that they are allowed to see layout tags. Both layout tags and format are considered to mean the layout of the source file. Out of those thirteen subjects, two people mentioned that, although they can see the tags, they do not know how to interpret them. Three people answered that it is important for them to see the source file or a preview of the translation, because that would allow seeing the source format. In particular, they mentioned TagEditor and Translator Workbench, which are integrated with Microsoft Word, and enable seeing the source layout.

Four people said that they are able to see comments and three that they are able to see glossary information.

3. When you translate documentation/online help files, what is the most usual file format you get?

According to the replies, the most used formats are XML, HTML and DOC, with fourteen, thirteen and twelve answers respectively. This is significant because it shows that markup languages are the most used file format for documentation/ online help.

It is also relevant mentioning that four people use XLIFF and five RTF[12], which are both interchange formats. As it was mentioned before, translators and translation agencies use a number of tools with different file requirements; interchange formats such as XLIFF and RTF are usually compatible with most tools.

Six people answered that they use TTX format, but these answers are not valid because TTX (Trados Tag) is a file format only used by SDL Trados TagEditor. TagEditor is an application used to translate markup languages files, such as HTML, XML and XLIFF, which are then converted into TTX to be processed by TagEditor. Therefore, TTX is a format only used with TagEditor, but it is not a standalone format.

4. If you translate XLIFF files, which XLIFF editor do you use? Do you think XLIFF editors allow you to see more information than other applications about the text or the project?

Most subjects replied that they use Translation Workspace as an XLIFF editor (eight subjects), which is due to the fact that many of the people that received this questionnaire work for Lionbridge as a freelance translator, and they are required to use that tool. Three subjects answered that they use SDL Trados.

In terms of the information seen with XLIFF, it is significant that six people do not seem to have observed any improvements in the way layout and information is displayed by XLIFF.

Finally, it is also relevant to mention that fourteen people have never worked with XLIFF.

5. When you translate documentation/ online help files, what kind of information do you usually miss?

Eight people answered that they usually miss context. This is a serious problem for translators, which can only be solved by project managers or clients providing information about the project.

[12] The Rich Text Format (often abbreviated RTF) is a proprietary document file format with published specification developed by Microsoft Corporation since 1987 for Microsoft products and for cross-platform document interchange.

It is also necessary that the translation tool enables recording such information so that the translator can easily access it.

Five people miss being able to see layout/ format information, and two people would like to understand what tags and placeholders stand for. In addition, three people say that as long as they can see the source file, they do not miss anything. Therefore, for nine people in total, it is important to see the layout of the source file.

Finally, project information, glossary, style guide and screenshots are also considered necessary for the translation of the files.

4.4. Proposal

After having analysed the way that XLIFF editors treat and filter the layout of source files (OpenDocument files) and after having identified the flaws of these tools from a translator´s point of view, this section offers a proposal of an alternative and additional display of information.

This proposal is based on what is possible to do according to the XLIFF specifications of XLIFF version 1.2, as published by OASIS[13]. Suggesting a complete new way of displaying layout information would not be realistic as it would require a very high technical knowledge.

XLIFF editors take an implementation (a version) of the XLIFF specifications to create filters for XLIFF files. Likewise, XLIFF converters are based on an implementation of XLIFF to convert source files into XLIFF files. This implies that some source information can be lost when files are filtered by XLIFF converters or XLIFF editors; although that depends on the implementation defined by developers.

The latest specifications of XLIFF offer a variety of possibilities on how to display information, not only in terms of layout, but in terms of project information, stages, etc.

As it was mentioned in the "XLIFF" section, the XLIFF document is composed of one or more sections, each enclosed within a <file> element. The <file> element consists of a <header>

[13] http://docs.oasis-open.org/xliff/v1.2/os/xliff-core.html

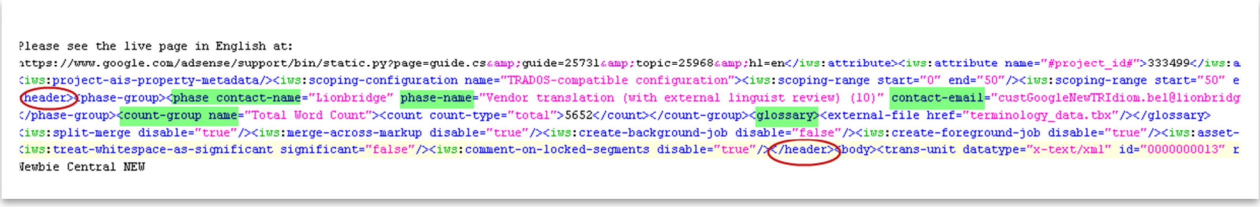
element, which contains metadata about the <file>, and a <body> element, which contains the extracted translatable data from the <file>. The translatable data within <trans-unit> elements is organised into <source> and <target> paired elements. In addition, XLIFF provides the ability to maintain information about the processing of the file via the <phase> element. The <source> element may contain inline elements that either remove the codes from the source (<g>, <x/>, <bx/>, <ex/>) or mask off codes left inline (<bpt>, <ept>, <sub>, <it>, <ph>).

This proposal focuses around two main blocks: information in the <header> element and information in the <body> element.

The <header> includes metadata information about the file, such as glossaries, translation memories and comments.

The Case Study of this research project has been carried out mainly using documents that are not real translation projects due to confidentiality issues; therefore, it has not been possible to evaluate thoroughly how XLIFF editors treat information about the project (word count, resources, etc). However, about seventeen real projects have been analysed and show that the <header> information can be seen with the XML or text editors, but not with XLIFF editors. These projects have not been used as examples in the Case Study because of confidentiality issues and because the format of the source file was not OpenDocument. Remember that this research project has focused solely on OpenDocument files.

For instance, figure 71 shows information of the header regarding: phase contact-name (contact for that specific translation stage), phase name (translation stage, i.e. translation, review), contact email, glossary information, count-group name (file word count).



```
Please see the live page in English at:
https://www.google.com/adsense/support/bin/static.py?page=guide.cs&amp;guide=25731&amp;topic=25968&amp;hl=en</iws:attribute><iws:attribute name="#project_id#">333499</iws:a
<iws:project-ais-property-metadata/><iws:scoping-configuration name="TRADOS-compatible configuration"><iws:scoping-range start="0" end="50"/><iws:scoping-range start="50" e
<header><phase-group><phase contact-name="Lionbridge" phase-name="Vendor translation (with external linguist review) (10)" contact-email="custGoogleNewTRIdiom.bel@lionbridg
</phase-group><count-group name="Total Word Count"><count count-type="total">5652</count></count-group><glossary><external-file href="terminology_data.tbx"/></glossary>
<iws:split-merge disable="true"/><iws:merge-across-markup disable="true"/><iws:create-background-job disable="false"/><iws:create-foreground-job disable="true"/><iws:asset-
<iws:treat-whitespace-as-significant significant="false"/><iws:comment-on-locked-segments disable="true"/></header><body><trans-unit datatype="x-text/xml" id="0000000013" r
Newbie Central NEW
```

Figure 68 - Header with jEdit

This information cannot be seen when opening the file with any of the three analysed XLIFF editors, which means that the XLIFF editors do not filter the header information accurately.

XLIFF editors should incorporate the possibility of portraying this information in the filtered file. It is extremely relevant for the translator to know as much information about the project as possible.

In terms of the <body>, there are several elements that have not been found in the corpora evaluated in the Case Study and that would be very valuable for translators, such as context. In addition, there are some elements that should be displayed in a different way with XLIFF editors.

a) Context

Contextual information for the localisation process can be provided by the <context> element, which is placed at the <trans-unit> or <alt-trans> level. It is vital for the translator to see context about the segment being translated, and, therefore, XLIFF editors should have a functionality that enables it. A given XLIFF editor could provide the information directly in the text with a <note> tag, or through other functionalities outside the text itself, such as segment comments or messages. There have not been examples of <context> elements in the Case Study because the files were not real translation projects, but also because, unfortunately, developers do not tend to include context in the source texts.

b) Hypertext reference

The href attribute (Hypertext reference) indicates the location of the file or the URL for an <external-file> element. The example below indicates a file on a local drive:

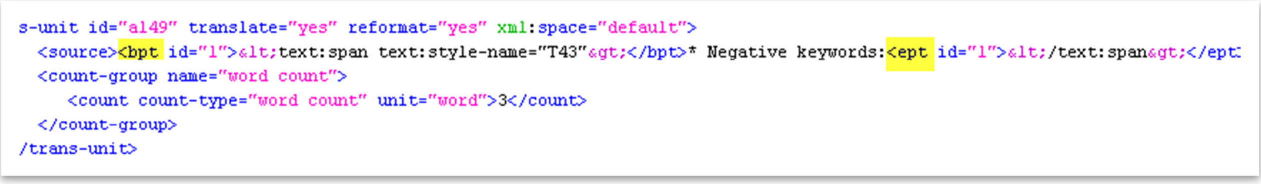
href=file:///C:/MyFolder/MyProject/MyFile.htm

The Case Study shows that images are not displayed by some XLIFF editors. Using a href attribute would be a solution to show the translator that there is a picture embedded.

c) Inline elements

Elements <bpt> and <ept> appear widely in the XLIFF files, but the translator is unable to identify the format of the segment because these inline elements just delimit the beginning of a paired sequence of native codes (see figure 75). In order to see the format of the document, these elements should have a “ctype” attribute, which specifies the type of code that is represented by the inline element; e.g. ctype="bold" means that the code represents a bolding text. None of the documents analysed with the XLIFF editors contained that attribute.

Note: In general, all inline elements should have attributes that specify the layout of the elements they delimit or replace. The XLIFF specifications allow tagging most layout information.



```
s-unit id="al49" translate="yes" reformat="yes" xml:space="default">
  <source><bpt id="1">&lt;text:span text:style-name="T43"&gt;</bpt>* Negative keywords:<ept id="1">&lt;/text:span&gt;</ept:
  <count-group name="word count">
     <count count-type="word count" unit="word">3</count>
  </count-group>
/trans-unit>
```

Figure 69 - BPT/EPT inline elements

d) Headings

Headings can be flagged with XLIFF using the element “restype” and its value “heading”. The attribute “restype” (resource type) indicates the resource type of the container element. It could appear in the <group> and <trans-unit> elements, such as <group restype='heading'>.

e) Lists

As with headings, lists could be flagged with XLIFF using the element “restype” and its value “list”. It can appear in the <group> and <trans-unit> elements.

f) Footnotes

They can be flagged with the value “fn” of the attribute “restype”.

As can be seen, XLIFF specifications offer a wide range of possibilities to mark layout and to include information about the project. The present proposal suggests two possible ways of displaying source text information. One option would be to flag this information with tags, as specified by XLIFF, but the downside would be that translators would have to know XLIFF to understand the tags. Another option would be that XLIFF editors filter XLIFF tags and convert them in intuitive information inserted in the text somehow so that translators understand it. For example, if there is a list of elements, the XLIFF editor could mark each segment with a "bullet-list" label.

As it can be inferred from this proposal, there are many options to identify information about the project and the layout of the source file in a more clear and visual way for translators. However, XLIFF editors are not able to use XLIFF to its full potential.

5. CONCLUSION AND FUTURE WORK

The initial hypothesis of this research project was that XLIFF editors do not provide all the text information that translators need because they are not taking advantage of all XLIFF features. The Case Study aimed at determining whether the XLIFF format adds value or hinders the translation process and whether XLIFF editor tools hide any relevant information for translators.

After having analysed the corpora, it was found that XLIFF editors do not display layout information in a clear and effective way. There are many examples in the Case Study about lost information, such as missing list elements and lack of text styles. This means that XLIFF editors should extend the XLIFF implementation they use. Furthermore, in those cases where XLIFF editors do provide information, it might still be lost if translators are not familiar with the markup language.

The XLIFF format potentially adds value to the translation process because it offers a wide range of helpful information to the translator, such as information about the layout of the text, the project, the localisation stage, etc. However, if XLIFF editor tools and converters use a limited implementation of the XLIFF format the translator will never benefit from that information. Therefore, XLIFF editor tools should adapt their functionalities to the variety of advantages that the XLIFF format offers.

XLIFF specifications could simplify their tag system to make it more intuitive for translators. As the questionnaire has proven, layout is important for translators but they do not always understand the tags, which means that either they need to be trained on how to interpret markup languages or translation tools need to have sophisticated filters that interpret tags in a self-explanatory way for translators.

As it was mentioned before, technology has shaped the translation world both in terms of translation tools and in terms of the translation needs. Technology translation is a very large market in the translation industry, which has introduced a new way of understanding markup languages. In addition, translation tools have become essential for the professional translation work. Consequently, it is vital to evaluate the new standards that are introduced in the industry to improve the work of translators, such as XLIFF. Unfortunately, although people have high hopes for its success, it does not seem to meet the high standards that it initially had.

This research project has aimed to flag the need of providing translators with the best tools possible to produce accurate and skillful translations. The Case Study has shown that XLIFF editors do not use XLIFF format to its full extent and that their implementation of XLIFF is far from being useful for the translator.

Software and documentation translators often complain about the fact that they translate blindly. Software translation is a very complex world that has not been analysed in this research project. However, the Case Study has proven that translators' complaints are actually well-founded, because the implementations of XLIFF editors do not seem to filter the information of the source text. It is of paramount importance that developers are aware of the needs of the translators to develop the right tools.

This research project has not used a very high number of samples for the Case Study, although a higher number of files were evaluated to understand XLIFF. It has focused on a few source texts that were rich-formatted and could serve the purposes of the study. However, in future research projects a higher amount of samples will be used to prove the conclusion. In addition, as it was stated earlier, not all samples were real translation projects. In an ideal situation, the samples would have been actual source texts for translation but it is somewhat difficult to have access to real cases due to confidentiality agreements that translators sign.

The source file format that was chosen for this research project is the OpenDocument. In the future, it would be interesting to expand the range of source files to other formats, such as HTML, DOC or XML, as they might be differently filtered by XLIFF editors. This would add more value to the conclusions reached in this research project.

In addition, it would also be interesting to expand the number of XLIFF editors used to evaluate the corpora and see whether there are tools that filter XLIFF format better for translation.

Finally, translators that work with XLIFF files and with other markup languages must raise their voices to those who build tools for them. Although the XLIFF format is a potentially valuable format that has been supported by experienced people in the translation world, it is not working as well as it could because many tools are not sophisticated enough to treat this format appropriately. Unfortunately, translators are at the end of the supply chain in the translation industry and their interests are not always prioritised. As a result, translators have to translate whatever comes to their hands and if it is an unintelligible XLIFF file, they might end up being “lost in XLIFF translation”.

6. REFERENCE LIST AND BIBLIOGRAPHY

BOOKS AND JOURNALS

Austermühl, F. 2001. *Electronic Tools for Translators*. Manchester: St Jerome.

Baker, M., 2001. *Routledge Enciclopedia of Translation Studies*. 2nd ed. New York: Routledge.

Bowker, L., 2002. *Computer-Aided Translation Technology: a practical introduction*. 1st ed. Canada: University of Ottawa Press.

Esselink, B., 2000. *A Practical Guide to Localization*. 1st ed. Amsterdam/ Philadelphia: John Benjamins B.V.

Hutchins, J. 1998. The Origins of the Translator´s workstation. *Machine Translation*, 13 (4), 287-307.

Kelly, D., 2005. *A Handbook for translator trainers*. 1st ed. Manchester: St. Jerome Publishing.

Melby, A.K.: 1992, The translator workstation. In J. Newton, ed. *Computers in translation: a practical appraisal.* Routledge, London, pp. 147-165.

Musciano, C., & Kennedy, B,. 2007. *XML and XHTML: The Definitive Guide*. 6th ed. Sebastopol: O´Reilly.

Oliver, A., Moré, J., & Climent, S., 2007. *Traducción y Tecnologías*. 1st ed. Barcelona: Editorial UOC.

Pym, A., 2004. *The Moving Text: Localization, translation, and distribution*. Amsterdam/ Philadelphia: John Benjamins Publishing Company.

Quah, C. K. 2006. *Translation and Technology*. New York: Palgrave Macmillan.

Somers, H. 2003. *Computers and Translation. A translator's guide*. Amsterdam/Philadelphia: John Benjamins. pp:31-48.

Sprung, R.C. Ed., & Jaroniec, S., 2000. *Translating Into Success: Cutting-edge strategies for going multilingual in a global age.* Amsterdam/ Philadelphia: John Benjamins Publishing Company.

ELECTRONIC RESOURCES

Alcina, A., 2008. *Translation Technologies: Scope, tools and resources.* [pdf] Amsterdam: John Benjamins Publish company. Available at: http://www.benjamins.com/jbp/series/Target/20-1/art/05alc.pdf [Accessed 15 May 2010].

Corrigan, J., Foster, T. 2010. *XLIFF: An Aid to Localization.* [Online] Available at: http://developers.sun.com/dev/gadc/technicalpublications/articles/xliff.html. [Accessed 02 July 2010]

Corte Fernández, N., 2002. *Localización e Internacionalización de sitios web.* Revista Tradumática. 1. Available at: http://www.fti.uab.es/tradumatica/revista/articles/ncorte/art.htm. [Accessed 02 July 2010]

Fernández García, J.R., 2006. Cambios de Herramientas. *Linux Magazine*, [Online]. 22, p. 5. Available at: http://www.linux-magazine.es/issue/22/Educacion.pdf. [Accessed 12 June 2010].

Ford, J.L., 2010. *HTML, XHTML and CSS for the Absolute Beginner.* [Online] Boston: Cengage Learning. Available at:
http://0-proquest.safaribooksonline.com.fama.us.es/9781435454231. [Accessed 18 April 2010].

Free Software Foundation, 2010. *What is free software?* [Online] Available at: http://www.fsf.org/about. [Accessed 1 August 2010]

Java.net, 2010. Open Language Tools. [Online] Available at: https://open-language-tools.dev.java.net/. [Accessed 1 July 2010]

jEdit, 2010. About jEdit. [Online] Available at: http://www.jedit.org/. [Accessed 1 July 2010]

LISA, 2010. *About LISA* . [Online] Available at: http://www.lisa.org/. [Accessed 15 July 2010]

LISA, 2010. *XLIFF version 1.2*. [Online] Available at: http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html/. [Accessed 2 July 2010]

Nuñez Piñeiro, O., 2006. Summary of a discussion on: What is XML and how do we teach it? In: A. Pym, A. Perekrestenko, P. Starink, eds. 2006. *Translation Technology and its Teaching (with much mention of localization).* [pdf] Tarragona: Intercultural Studies Group URV. 65-66. Available at: http://isg.urv.es/publicity/isg/publications/technology_2006/index.htm [Accessed 9 April 2010].

OASIS, 2010. *About Oasis*. [Online] Available at: http://www.oasis-open.org/who/. [Accessed 5 July 2010]

OASIS, 2010. *OASIS XML Localisation Interchange File Format TC*. [Online] Available at: http://www.oasis-open.org/committees/xliff/faq.php/. [Accessed 5 June 2010]

OASIS, 2010. *XLIFF Version 1.2*. [Online] Available at: http://docs.oasis-open.org/xliff/xliff-core/xliff-core.html. [Accessed 4 June 2010]

Opentag.com, 2009. *XLIFF*. [Online] Available at: http://www.opentag.com/xliff.htm. [Accessed 30 June 2010]

ORACLE, 2010. *XLIFF: An aid to localization*. . [Online] Available at: http://developers.sun.com/dev/gadc/technicalpublications/articles/xliff.html. [Accessed 30 May 2010]

Pym, A., Perekrestenko A., Starink, P., 2006. *Translation Technology and its Teaching (with much mention of localization).* [pdf] Tarragona: Intercultural Studies Group URV. Available at: http://isg.urv.es/publicity/isg/publications/technology_2006/index.htm [Accessed 9 April 2010].

Raya, R., 2010. *XML in localisation: Use XLIFF to translate documents*. [Online] Available at: http://www.maxprograms.com/articles/xliff.html. [Accessed 20 May 2010]

Translation Solutions, LTD. *How to translate XML?* [Online] Available at: http://www.your-translations.com/XML_translation_basics_1.php. [Accessed 10 May 2010]

W3 Schools, 2003. *XML Tutorial*. [Online] Available at: http://www.w3schools.com/xml/default.asp. [Accessed 20 May 2010].