Translation Challenges in the Localization of Web Applications

Elena de la Cova
mecovmor@upo.es
Universidad Pablo de Olavide, Seville

Recibido: 20/06/2016 | Revisado: 20/07/2016 | Aceptado: 08/11/2016

Abstract

This preliminary study aims at exploring the nature of challenges that translators face when they take on a localization project of a web application. Taking into account that localization is an activity constrained by time, process and economic resources, translators need to make use of their full skill set to overcome the various challenges imposed by the source text and the localization process itself.

For the purpose of this study, an *ad hoc* monolingual English corpus composed of the user interface strings of web applications has been used. Since multiple types of challenges are found in a localization project of this nature, this paper focuses on those related to internationalization practices and to constraints imposed by the translation memory segmentation process.

Although localization is a mature field and a great deal of guidelines and best practices is available for content creators and tool developers, as found in this qualitative study, localizers can still suffer the consequences of deficient internationalization practices and non-ergonomic translation tools.

Keywords: localization, web applications, challenges, internationalization, translation memory systems

Resumen

Desafíos de traducción en la localización de aplicaciones web

Este estudio preliminar tiene como objetivo explorar la naturaleza de los desafíos a los que los traductores se enfrentan cuando se embarcan en un proyecto de localización de una aplicación web. Dado que la localización es una actividad condicionada por tiempo, procesos y recursos económicos, los traductores tienen que poner en marcha todas sus competencias para superar los muchos desafíos impuestos por el texto fuente y por el proceso de localización en sí.

En este estudio, se ha utilizado un corpus monolingüe *ad hoc* en inglés compuesto por mensajes de la interfaz de aplicaciones web. Puesto que en este tipo de proyecto de localización existen diferentes tipos de desafíos, este artículo se centra en aquellos relacionados con las prácticas de internacionalización y con la segmentación de las memorias de traducción.

A pesar de que la localización es un ámbito de considerable madurez y de que los creadores de contenido y los desarrolladores de herramientas tienen a su disposición un gran abanico de instrucciones y directrices, en el presente estudio cualitativo se concluye que los localiza-

dores siguen sufriendo las consecuencias de prácticas de internacionalización deficientes y de herramientas de traducción poco ergonómicas.

Palabras clave: localización, aplicaciones web, desafíos, internacionalización, sistemas de memoria de traducción

1. Introduction

The localization industry stresses the importance of cultural adaptation of products and services for their success in target markets (DePalma 2006; Schäler 2007). Unfortunately, the impact of localization on a given business often becomes visible in society only when its results are poor. There are a number of authors that have explored translation mistakes in localization. Diéguez and Lazo (2011) analyze successes and errors found in localized websites and categorize the translation techniques followed according to Hurtado Albir's classification (2001: 266-271). Jiménez-Crespo has written extensively about the corporate website as a text genre from different perspectives (2008; 2009a; 2009b; 2011a; 2011b; 2012; 2013), and has proposed a holistic error typology based on a monolingual comparable corpus of Spanish original and localized corporate websites. His research provides an additional empirical methodology to current quality assessment systems in the localization industry, which lack a comprehensive outlook and could benefit from aspects identified in his error typology such as pragmatic errors (Jiménez-Crespo 2011a). Additionally, the quality of website localization has also caught the attention of researchers like Medina and Ramírez (2015), who have evaluated the quality of the "Products" section from SME corporate websites.

Pym (2010) highlights the lack of studies about translation mistakes in localization, despite the severe consequences these mistakes have on businesses. A case to illustrate this took place in 2015, when the well-known Spanish retailer Zara (from the Inditex retail group) published in their German online shop a new model of sandals called "*dreifarbige Sklaven Sandalen*" (direct translation from the Spanish "*sandalias de esclava*", in English "three-colored slave sandals"). According to the RAE Spanish dictionary (Royal Spanish Academy), the word "slave" ("*esclava*") in Spanish can also refer to a kind of simple bracelet design. Therefore, Zara apparently used the word "slave" in an attempt to compare their sandal design to that of the bracelet. However, translating "*sandalias de esclava*" literally into "*dreifarbige Sklaven Sandalen*" immediately triggered the offense from the German audience because "*Sklaven*" does not have the same meaning as in Spanish (type of bracelet). This error went viral on social media (as can be seen on Twitter under the hashtag #sklavensandalen), and Zara had to issue a press release explaining that they do not promote slavery and it was a translation mistake. It was, nevertheless, too late and criticism spread to other business areas, such as the working conditions of their labor in developing countries.

Although they apologized and removed the word "slave" from the design name, the company reported a decrease in their sales in Germany the days following the "translation mistake" (lavanguardia.com, 2015). Germany is one of Zara's biggest markets and therefore this issue had severe consequences.

In order to provide content with optimal quality and adequate localization, translators need to have the right competences and tools. Conversely, the reality of the translation and - especially - the localization industry does not always allow translators to make the most out of their translation skills. Deadlines are usually tight, translation instructions might not be very clear and the translation environment limits the way translators work. According to Alonso (2016), professional translation environments might currently be ruled by conflict, opacity and mistrust between translators and direct clients or managers, due to the computer-mediated nature of communication and of translation production networks (Abdallah and Koskinen 2007). Localization is often the last stage before content is published (Herrmann and Sachse 2005: 69) and, depending on the process, it may happen that localizers do not have the freedom or the possibility to make the changes necessary to adequately adapt the content to their markets. Additionally, communication barriers between authors and localizers frequently prevent feedback to flow along the globalization process, which would be greatly beneficial for content creation (Fenstermacher 2006: 82).

This paper aims at exploring the challenges that localizers face when they are exposed to the localization of a web application project. Given that there can be a myriad of challenges in this process, this qualitative study focuses on how the "internationalized text" (Pym 2004), or the lack of proper internationalization, can affect the process. In addition, since the intersection between translation and technology is inherent in localization (Jiménez-Crespo 2010) and it affects how localizers work, the translation environment has been taken into consideration in the present study. In particular, the translation memory segmentation process has been observed as a potential challenge that translators have to overcome.

In order to identify translation challenges arising from the localization of web applications, an English monolingual corpus of web applications has been analyzed. In this paper the concept of *challenge* is essentially understood as the notion of *translation problem*, which has been widely analyzed by different authors in Translation Studies, such as Wilss (1982), Lörscher (1991), Kiraly (1995), Presas (1996), Nord (2005), Toury (2010) or PACTE group (2011), among others. Although exploring this complex concept in detail is not the purpose of this paper, a definition of *translation problem* will be provided to frame the perspective of the study:

> A translation problem is an objective (or inter-subjective) transfer task which every translator (irrespective of their level of competence and of the technical working conditions) has to solve during a particular translation process.
>
> (Nord 2005: 166)

As will be seen in this empirical study, many of the translation challenges can be solved more or less successfully depending on various factors but they are still considered translation problems. Therefore, according to the perspective of this study, a translation problem is viewed as such regardless of whether it can be solved or not and regardless of the experience or "level of competence" of the translator, who might have been exposed to a particular challenge in the past and learnt to overcome it. That, however, does not make the challenge itself less of a challenge; it just makes it easier to sort out for a given translator.

This paper intends to depict some of the challenges that localizers face when they take on the localization of a web application which, if not dealt with adequately, could cause localization and translation errors. Therefore, this qualitative study should be seen as an attempt to explore the previous step to the localized outcome (the translation stage). Unfortunately, due to non-disclosure agreements signed by translators - among other factors - researchers might find it difficult to observe and reflect on a paper what translation challenges localizers deal with in professional projects and how they solve them (Rojo 2013: 71). This could be seen as a methodological constraint imposed by the localization industry processes. In this sense, the present qualitative study aims at exploring the localization of web applications at the translation stage. In order to do that, an empirical approach based on a corpus - as recommended by Olohan (2004) and Gamero (2001) cited in Jiménez-Crespo (2011a: 319) - together with the researcher's intuition and professional experience in this field have been used to determine what constitutes a localization challenge.

Since this study focuses on the localization of web applications, section 2 covers basic concepts related to localization. Section 3 presents the topic and the theoretical background behind each of the challenges identified. Section 4 explains the methodology followed and illustrates those translation challenges with examples from the analyzed corpus. Finally, section 5 highlights the conclusions of the study.

2. Localization

The localization industry was born in the eighties as a result of the technology boom and the arrival of desktop computers to the general public (Castells 2005; Esselink 2003). As technology became more and more present in the daily life of ordinary people, the demand for computer programs in local languages increased. A new translation profile was required; mainly a professional translator with technical skills who could embark on new and technologically advanced projects. As Esselink and Cadieux (2002: 2) put it: "when people started translating software, some of the changes required were not, strictly speaking, translation: changes to character encodings, date and time formats, sorting rules, etc." Due to the need of specialized teams to translate or localize software, many high-tech companies outsourced their localization projects to translation vendors in Europe, mostly in Ireland, which offered certain tax benefits to multinational high-tech companies based in Ireland (Esselink 2003). Mi-

crosoft and Oracle, among many others, had their European headquarters in Dublin and, as their need for localization services grew, the number of localization vendors established in Ireland grew as well (e.g. INK, later Lionbridge).

The term *localization* has been traditionally defined by LISA (Localization Industry Standards Association)[1] as:

> 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.
>
> (LISA 1998: 3)

Although this definition does not specify what kind of products are typically localized, in a broad sense localization refers to adapting some kind of technology or digital product to a given target culture from several perspectives. As GALA[2] (Globalization and Localization Association) explains:

> True localization considers language, culture, customs, technical and other characteristics of the target locale. While it frequently involves changes to the software writing system, it may also change the keyboard usage, fonts, date, time and monetary formats. Graphics, colors and sound effects also need to be culturally appropriate.
>
> (GALA 2013)

Thus, when a new piece of software is designed to be launched in ten different international markets, localizers step in to adapt it culturally and linguistically to make sure that its users feel that it has been originally written and designed in their own language. And yet, that goal is not so easy to achieve if the product has not been properly internationalized, as it will be shown later on.

According to Jiménez-Crespo (2011: 4), there are five types of localization: software, websites, video games, small devices and multimedia content. However, there are two primary and distinctive localization processes: software localization and website localization or, as Mata (2005: 190) suggests, computer program localization (*Localización de productos informáticos* or LPI) and web content localization (*Localización de contenido web* or LCW). The former refers mainly to software products but also to mobile apps, video games and software documentation and the latter applies to products hosted in or created by technologies traditionally linked to the World Wide Web.

Software localization has been defined by Esselink (2000: 57) as "the translation of all graphical user interface (GUI) components of a software application, such as dialog boxes, menus, and error or status messages displayed on screen". Software localization involves working with or working around elements from the programming environment or processes, like placeholders, control codes, and concatenated strings. Furthermore, software localization implies taking into account aspects such as space

restriction, formats conversion, lack of context, bug fixing and image localization (Alonso and De la Cova 2013).

Space restriction is a particularly challenging aspect in software localization because text space in a software application is limited (even more in mobile applications). Therefore, the possibility of text expansion or restriction should be considered. Languages do not take the same space or character length for a given text. For example, Spanish uses 17% more characters than English for the same text (Oliver, Moré, Climent 2008: 178).

In addition, image files and multimedia content are increasingly part of software localization projects but their localization is seen by companies as quite an endeavor more from the economic and organizational perspective than from a technical point of view (Mata 2009).

In terms of website localization, although several authors have defined it (Corte 2002; Cronin 2003; Sandrini 2005; Pym 2011a), for the purpose of this study, the following definition will be used:

> A complex communicative, cognitive, textual and technological process by which interactive digital texts are modified to be used in different linguistics and sociocultural contexts, guided by the expectations of the target audience and the specifications and degree requested by initiators.
>
> (Jiménez-Crespo 2013: 20)

This definition stresses the importance of the target audience and the specifications requested by the client or the product manufacturer, which is relevant to this study because those specifications can greatly impact the translation process and, to a certain extent, limit it.

Although software and website localization still entail different processes and requirements, the difference between them are starting to fade out. Web applications share some features of client-server traditional programs but are generated dynamically using the web browser as a client. As Yu and Offutt (2002: 5) explain, web applications generally work like this: "a client first retrieves information from servers in the form of HTML files, then sends requests to servers, and finally expects replies as HTML files."

It is therefore in this workflow of content where localizers intervene and where – as will be explained below - they face a number of challenges posed by the internationalization aspects of the text, as well as by translation tools themselves.

The situation of these challenges can be illustrated by the "dialect of resistance and accommodation" (Pickering 1995 in Olohan 2011: 344). According to this "dialect", resistance is offered by an entity, for example, technology; and accommodation is seen as the human response to that resistance. Olohan (2011: 342-355) applied this sociological theory to translation studies and, through the analysis of translators' posts on a TM tool online forum, she identified the dialect of resistance and accommodation

("dance of agency") in the interaction between the TM tool and the translators. In the wider context of localization processes, the internationalized text and technological barriers could provoke resistance among translators; whereas accommodation would take place when translators, forced by their *skopos* - together with tight deadlines and client expectations as depicted by Alonso (2016) - work around these challenges and end up aiming to produce an acceptable output.

3. Challenges in the localization of web applications

Although web applications can cause a variety of challenges for translators, due to space limitations, this study does not cover all of them but instead focuses on internationalization related problems. In particular, it discusses to what extent internationalization deficient practices (both technical and language related) leave certain traces in source texts that constrain translation activity. In addition, since localizers generally work with translation memory systems, challenges related to how the source text is segmented and presented to translators have also been taken into account.

Before moving to the empirical analysis of our corpus, we will review the existing literature pertaining to our object of study, i.e. localization challenges related to internationalization factors and segmentation processes. More particularly we focus on: i) challenges caused by software internationalization deficient practices, ii) challenges resulting from inadequate content internationalization, and iii) challenges due to the rigidity of the segmentation of translation memories.

3.1. Software internationalization

The internationalization process takes place during product development in order to enable localization:

> Internationalization (I18n) 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.
>
> (LISA 1998: 3)

Software internationalization is critical for the correct localization of a product, and it comprises two main activities: 1. Separating text from the development code (so that it can be easily translated in an independent resource), and 2. Enabling the code to support linguistic and cultural adaptation (Esselink 2000; McKethan and White 2005; Schäler 2007; Massion 2011).

In terms of the first activity, if the text and the source code are not appropriately separated there can be such complex and costly localization problems that localization might even be canceled. Therefore, programmers should separate user interface text and source code according to the set guidelines of the development environment

(Herrmann and Sachse 2005: 47). Although the separation of text for translation is the right practice to follow, it can trigger other sorts of challenges for localizers (see sections *User interface text separation* and *Ambiguity*).

With regards to the second activity, when developing a piece of software, the engineer should program the code in such a way that it will be able to display bi-directional text (left-to-right or right-to-left), e.g. to handle its localization in Hebrew. Additionally, it should allow international characters, such as Japanese *kanjis*, if it is going to be localized in that language, for which they need to make sure that character encoding is right. Therefore, it is paramount to decide whether specific software is going to be localized further down the line or not to avoid unnecessary problems and costs after the software has already been developed (Herrmann and Sachse 2005; Asnes 2007; Horvath 2007; Yang 2007).

In addition, there are certain international practices to be considered when developing software so that its localization can be carried out appropriately and without causing problems for translators. Most leading technology manufacturers provide detailed guidelines on how to internationalize their products correctly, such as Microsoft, Oracle or Google, as well as web technology contributors as the W3C, which offers broad web internationalization guidance in its "W3C Internationalization (I18n) Activity"[3] website.

Horvath (2007: 3) suggests a list of ten internationalization tips to overcome most common internationalization challenges. They range from using Unicode functions and methods to preparing a translation kit that contains all necessary information for translators. Among those ten tips, two are especially relevant for this study: user interface text separation and avoiding ambiguity.

Although internationalization is a widely documented process and there are numerous best practices guidelines available for publishers to use, localization problems can persist if those are not implemented. As Dunne (2006: 6) points out:

> The relative success of translation and localization efforts depends to a great extent on the successful implementation of internationalization strategies. However, these in turn depend on an enterprise level commitment to globalization strategies.

Moreover, the recent Multidimensional Quality Metrics (MQM) framework, which defines quality metrics for translation quality assessment, includes internationalization as one of the main dimensions of the ten identified:

> Internationalization covers areas related to the preparation of the source content for subsequent translation or localization. Internationalization issues may be detected through problems found in the target (particularly from those included in Locale convention (locale-convention), but an Internationalization audit is generally conducted separately from a general assessment of translation quality. (Lommel, Burchardt and Uszkoreit 2016)

In sum, these details reveal that although internationalization is widely implemented for content globalization and best practices documentation are available, it is still a cause of concern in the field in terms of quality assessment, and also in terms of costs inefficiencies (LISA 2007).

3.1.1. User interface text separation

As mentioned earlier, separation of text from code is critical for the correct and efficient localization of a product. In fact, the "Mozilla Localizability Guidelines" mention as their first recommendation "Don't mix data and code" (Mozilla 1998) referring essentially to avoiding hard-coded strings and separating text in a resource file. A resource script file is "a text file containing descriptions of resources from which the resource compiler creates a binary resource file" (Esselink 2000: 473). *Hard-coding* is a coding process that involves "the embedding of translatable strings in the body of programming code rather than in separate resource files" (Esselink 2000: 469). To put it another way, when an element such as a name or a message is hard-coded, the translator does not have access to it in context (or at all); meaning that the linguistic context where it appears (sentence or TM segment) does not contain the element itself (which is embedded in the code). Several internationalization best practice guides (Mozilla 1998; Vine 2002; Horvath 2007; Yang 2007) advice against hard-coding because it can potentially cause translation quality issues, translation inconsistencies or the publication of outdated or untranslated content.

Similarly, engineers resort to different ways of using elements that work like placeholders but are not accessible to translators. Placeholders or variables are defined as "characters that are usually preceded by a percentage (%) sign and replaced by another word, value or string at application run-time" (Esselink 2000: 68). This is illustrated in the following example drawn from our corpus:

Are you sure you want to reset **'s password?**

On this example, the HTML element would be replaced at runtime by the username of the person whose password will be reset (i.e. Are you sure you want to reset Peter's password?). Wrapping text with tags provides a hook to an element most likely via JavaScript[4] (Roberts 2014). If the element was a placeholder, the translator would have access to it and would be able to move it where necessary depending on their language structure. For example, in Spanish the order of the possessive noun phrase elements would be reversed:

¿Seguro que deseas restablecer la contraseña de **?**

However, in this particular example, the element would most likely not be accessible in a TM environment and therefore the localizer would not be able to edit it. The sentence would be parsed in two different source segments (SS):

TM SS 1: **Are you sure you want to reset**
TM SS 2: **'s password?**

Although in this case the translator could tweak the phrasing to adjust to the Spanish language structure, the solution is far from ideal because the TM target segments (TS) would not be accurately stored (as will be described further in section 4.2.3.):

TS 1: ***¿Seguro que deseas restablecer la contraseña de***
TS 2: **?**

3.1.2. Ambiguity

The second internationalization deficient practice that Horvath (2007: 3) advises against is *ambiguity*, which results mostly from not having context in GUI texts (either linguistic or visual context).

Although separating text from source code is a fundamental internationalization practice, it means that the GUI text will be stored in a resource file or database (basically in a correct resource format) without context. That resource file then is converted to a file format that translators can use in their translation memory systems (e.g. XLIFF file).

> The source files are usually programming code files-e.g. instructions in the programming language C++- and resource files (with the extension .rc), which include user interface elements such as dialog boxes, which constitute the main elements to be translated. (Sandrini 2008: 170)

Therefore, if the software strings do not contain contextual information, they might be ambiguous and difficult to understand (Massion 2011; Horvath 2007). Unfortunately, this is too often the case (Safar and Machala 2010). Figure 1 is an example of a typical resource file where the text of the software is stored:

Figure 1. Resource file of WinHTTrack

```
    GROUPBOX        "On hold",IDC_STATIC_delay,20,122,225,37
    LTEXT           "Transfer scheduled for: (hh/mm/ss)",IDC_wait,25,132,215,8,WS_DISABLED
    EDITTEXT        IDC_hh,25,143,16,12,ES_AUTOHSCROLL | ES_NUMBER
    EDITTEXT        IDC_mm,42,143,16,12,ES_AUTOHSCROLL | ES_NUMBER
    EDITTEXT        IDC_ss,60,143,16,12,ES_AUTOHSCROLL | ES_NUMBER
    GROUPBOX        "",IDC_STATIC,7,167,242,31
    CONTROL         "Save settings only, do not launch download now.",IDC_select_save,
                    "Button",BS_AUTORADIOBUTTON | BS_MULTILINE,12,174,233,20
END

IDD_wizard DIALOG  0, 0, 190, 98
STYLE DS_SETFONT | WS_CAPTION | WS_SYSMENU
CAPTION "Wizard query"
FONT 8, "MS Sans Serif"
BEGIN
    EDITTEXT        IDC_reponse,106,58,77,12,ES_AUTOHSCROLL
    DEFPUSHBUTTON   "OK",IDOK,133,77,50,14
    LTEXT           "Your answer:",IDC_STATIC_answer,15,62,87,8
    LTEXT           "",IDC_question,7,7,176,47
END

IDD_inprogress DIALOG  0, 0, 361, 267
STYLE DS_SETFONT | WS_CHILD
FONT 8, "MS Sans Serif"
BEGIN
    LTEXT           "In progress:",IDC_STATIC_inprog,15,7,70,8
    LTEXT           "",IDC_inforun,85,7,265,8
    GROUPBOX        "Information",IDC_STATIC_informations,7,26,343,59
    LTEXT           "Bytes saved:",IDC_STATIC_bytes,15,40,65,8
    LTEXT           "",IDC_i0,85,40,110,8
    LTEXT           "Links scanned:",IDC_STATIC_scanned,199,40,66,8
    LTEXT           "",IDC_i2,274,40,72,8
    LTEXT           "Time:",IDC_STATIC_time,15,50,65,8
    LTEXT           "",IDC_i1,85,50,110,8
    LTEXT           "Files written:",IDC_STATIC_written,199,50,66,8
    LTEXT           "",IDC_i6,274,50,72,8
    LTEXT           "Transfer rate:",IDC_STATIC_trate,15,60,65,8
```

In the previous figure, the savvy localizer would be able to identify certain elements (buttons, menus, etc.) that would add some kind of context to their translation. Nevertheless, quite frequently, the file sent to translation would look very much like this, i.e. a list of strings without any contextualization or information of the nature of the user interface component:

Figure 2. GUI string list of WinHTTrack

```
Archivo  Edición  Formato  Ver  Ayuda
In progress:
In progress:
Follow external links
Follow external links
Test all links in pages
Test all links in pages
Try to ferret out all links
Try to ferret out all links
Download HTML files first (faster)
Download HTML files first (faster)
Choose local site structure
Choose local site structure
Set user-defined structure on disk
Set user-defined structure on disk
Use a cache for updates and retries
Use a cache for updates and retries
Do not update zero size or user-erased files
Do not update zero size or user-erased files
Create a Start Page
Create a Start Page
Create a word database of all html pages
Create a word database of all html pages
Build a complete RFC822 mail (MHT/EML) archive of the mirror
Build a complete RFC822 mail (MHT/EML) archive of the mirror
Create error logging and report files
Create error logging and report files
Generate DOS 8-3 filenames ONLY
Generate DOS 8-3 filenames ONLY
Generate ISO9660 filenames ONLY for CDROM medias
Generate ISO9660 filenames ONLY for CDROM medias
Do not create HTML error pages
```

Lack of context is one of the main challenges that translators face when localizing a piece of software (Esselink 2000; Ludwigsen, Williams and Polis 2009; Safar and Machala 2010; Sikes 2010; Jiménez-Crespo 2013; Dunne 2015; Edwards 2015). Am-

biguity can cause confusion, translation errors and, ultimately, it can affect the quality of the localized product negatively (Herrmann and Sachse 2005). Words in isolation can cause confusion, have different meanings depending on the context (Bernal Merino 2007: 31-32) and they can also create problems to Machine Translation systems when options are not appropriately tagged (Roturier and Lehmann 2009).

To avoid working with de-contextualized strings that show in the resource file without any textual order, developers should include a description field where they can enter comments in order to provide context and additional information for translators (Ludwigsen, Williams and Polis 2009).

As Edwards points out (2015: 20):

> (…) context is so critically important to our work because so much depends on conveying the right information across linguistic and cultural barriers. Without all the proper contextual cues and metadata, our jobs become extremely challenging — if not even impossible in some cases.

This process of localizing text without context or descriptions could be termed "blind localization" because translators work very often without knowing what they are actually translating.

3.2. Content internationalization

As described in the previous sections, the term *internationalization* mainly applies to the technical process of designing a software or digital product in such a way that it enables localization; alternatively, the term can also apply to writing content with an international audience in mind (Alonso and De la Cova 2013). This concept is broadly referred to as "writing for translation" (also "writing for an international audience" or "writing world ready content"), but to emphasize the importance of considering international aspects when writing content, in this paper the phrase "content internationalization" is the one primarily used. As Brown (2003: 4) states: "Writing for translation requires the technical writer to express complex concepts in terms clear enough to move easily from one language to another".

Basically, writing for translation involves thinking ahead and writing text with an international audience in mind when a product is going to launch globally. This is central in avoiding problems at the localization/translation stage, like unnecessary costs and quality issues (Hoft 1995; Amant 2005; Sichel 2009; Yang 2007). In addition, Keller (2011) points to resulting time inefficiencies when working with TM systems as a further problem.

In particular, unnecessary costs can be prevented to a large extent if writers take into account that their output will be sent for translation but, unfortunately, that does not happen as often as it should (Brown 2003; Sichel 2009), as will be shown in section 4.2.2. Localization is a more expensive task than writing; for example, manufacturers may want to take into consideration that "an average of $4 is spent on localiza-

tion for every $1 spent on technical documentation" (Brown 2003: 4). Therefore, for a business that values their ROI (return on investment), this data should make them rethink their writing process when localization enablement is needed.

Because every culture has different ways of conveying basic information, it seems sensible that technical writers are aware of them in order to avoid problems and to help their companies prosper in a world where online communication is essential (Amant 2001). However, as it is not realistic to expect that writers would learn in detail about global cultural differences, there is a considerable amount of guidelines and manuals to help them with that task. This is particularly relevant for technology companies that have a substantial international presence and whose revenues largely come from international markets. For instance, Microsoft, Google or Symantec have large teams of writers (also called technical writers, editors or communicators). As Amant (2000: 326) points out:

> English's status as an international language does not mean that English-speaking technical communicators can now rest on their linguistic laurels. Rather, technical communicators should be aware of important technical publications printed in languages other than English in order to get a more comprehensive picture of the developments in a certain field.

Text written for translation should avoid culture-specific and location-specific references (Kamerer 2003), as they can be difficult to convey in other languages. Visual representations can also be problematic if they are transferred to international markets, although they are a very effective means of communication if used in the right way (Amant 2005).

Apart from the fact that the text should be written with translation in mind, the contextual information helps translator learn about the subject matter and bridge the gap between cultural differences (Flint et al. 1999: 239).

As established above, cultural references are one of the key elements to consider when writing content for an international audience. Hoft (1995) refers to these elements as *international variables*: "International variables are the localizable elements. International variables identify superficial and deep cultural differences" (Hoft 1995: 19). Usually, these variables are related to politics, economy, religion, education, linguistics and technology.

In order to help writers create "localization friendly" content, companies such as Microsoft and Mozilla create their own internationalization guidelines. Similarly, there is a significant amount of scientific production related to writing for translation primarily in the Technical Communications field. Nonetheless, there is only a limited amount of scientific production arising from Translation Studies about the importance of designing technical content for a global audience. De la Cova (2014) suggests a categorization of guidelines for international writers, specifically designed for web content localization. These categories include cultural references (e.g. humor), time and localization references (e.g. date formats), clear writing (e.g. avoiding ambiguity),

world-ready wording (e.g. rethinking slang), acronyms (e.g. DIY), colors and images (e.g. cultural-rich images) and punctuation (e.g. number formats).

3.3. Rigidity of translation memories

The third translation challenge highlighted in this paper derives from the rigidity of translation memories (TMs).

> 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 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.
>
> (Esselink 2000: 362)

Born in the seventies and widely implemented in the nineties (Somers, 2003), TMs have evolved considerably in the last decades, becoming an essential tool for professional translators and other language services providers. TMs have changed and shaped the way translators work (Bowker 2002) and have brought benefits in terms of cost savings and productivity gains (Lagoudaki 2006; O'Hagan 2009). Consequently, there is a large volume of studies related to the impact of TMs on the translation process (Bowker 2005; Mossop 2006; Biau Gil and Pym 2006; Jiménez-Crespo 2009; Pym 2011b; Olohan 2011; Christensen and Schjoldager 2011; LeBlanc 2013; Alonso 2015; Alonso and Calvo 2015; Bundgaard, Christensen and Schjoldager 2016). Research topics range from functional aspects about TMs, to the way they alter the translation process or to their influence on cognitive processes.

LeBlanc (2013) has conducted a study with Canadian translation vendors to evaluate their perception of translation technology, in particular regarding TMs. The results show the main advantages and disadvantages of using this technology from their perspective. On the one hand, among the advantages, LeBlanc highlights that TMs help increase productivity; they improve consistency; they eliminate repetitive work; they are used as databases; and they have a pedagogical function. On the other hand, among the disadvantages, translators have argued that TMs change their relationship with the text (due to segmentation); they are a barrier to creativity; they make translators passive; they have an effect on their natural reflexes; they make beginner translators too dependent on them; they are sometimes polluted and contribute to error propagation; they have a demanding impact on their productivity requirements; they are used as rigid guidelines for using exact or full matches; and finally, they can potentially have a negative effect on translators satisfaction if misused (LeBlanc 2013: 6-10).

One of the main challenges detected in the present study comes from the process of *segmentation*. Segmentation is one of the five main processes of TM systems: text segmentation, alignment, indexing, search and retrieval (Lagoudaki, 2006: 3). It is de-

fined as the "division of text into translatable units, such as sentences or paragraphs" (Esselink 2000: 473) or "other more or less easily distinguishable text portions, such as titles" (Somers 2003: 34). When translators enter their translations in target segments, the source and the target segments form the so-called "translation unit".

Different authors have focused on segmentation and how it affects the translation process (Dragsted 2006; Colominas 2008). For instance, Colominas (2008: 352) affirms that adding sub-sentential segmentation (chunk segmentation) to TM technologies might enhance their recall (target-text proposal) and precision (usability).

Challenges related to segmentation were also mentioned by translators in LeBlanc's study, generally with regard to the inflexibility of the sentence-by-sentence approach, which they describe as too mechanical and unnatural and possibly leading to problems in the cohesion of the text (LeBlanc 2013: 9). According to Bass (2006: 76), segmentation is one of the technological barriers identified in his study as obstacles to quality in the localization industry. Another relevant study (Jiménez-Crespo 2010: 203) comparing original and TM translated texts from the same genre shows that translated text replicates source superstructure, which would imply that TMs do restrain and constrain translators.

4. Empirical study

4.1. Methodology

This is a qualitative and descriptive study based on a single monolingual English *ad hoc* corpus composed of the GUI strings of three web applications from the cloud storage and social media field. The present study is a preliminary work related to a wider PhD research project (De la Cova, in progress) in which localization problems are being identified in a single monolingual English corpus and categorized following a proposed set of criteria.

According to EAGLES (1996), a corpus is "a collection of pieces of language that are selected and ordered according to explicit linguistic criteria in order to be used as a sample of the language". The corpus analyzed in this study is a single monolingual English *ad-hoc corpus* of 101,263 source words (18,911 strings), composed of the GUI strings of three web applications. For illustrative purposes, we have provided translations into Spanish when applicable.

These web applications have been chosen according to representativeness criteria in the application category they fall under, as they are some of the most used cloud storage and social media services of the last years.

In terms of corpus size, it satisfies the research purpose of this preliminary study. Although there does not seem to be an agreed optimal corpus size among researchers (McEnery and Wilson 2001, Sinclair 2002, Corpas and Seghiri 2006, Reppen 2010), some authors, such as Kennedy (1998 in Olohan 2004: 46), agree that a big corpus is not necessarily better than a smaller one and that the aim of the research should

shape the corpus design criteria. Similarly, Bowker and Pearson (2002: 45) state that "there are no hard and fast rules that can be followed to determine the ideal size of a corpus", and the researcher should determine what is the appropriate size depending on the purpose of the project and the availability of data. These authors suggest 25,000 words as an initial goal for a corpus size, which could be expanded.

Given the nature of the texts and the purpose of this analysis, a manual analysis seemed the most appropriate approach. As argued by Sinclair (2000 in Olohan 2004: 62-63), in qualitative studies corpus tools might be less useful than in automated analyses, where annotated corpus seem more suitable. Moreover, since the aim of the study was to observe the source text as translators do when they undertake a localization project in order to identify potential challenges, we concluded that using a TM tool would be a feasible approach. As a result, our corpus was processed in a TM tool, and segments were analyzed in the TM environment.

For the present study, the corpus has been processed in Smartling[5] Translation Management System because it is a tool specifically designed for websites and dynamic web applications. Smartling TM follows regular industry segmentation practices and includes standard TM features. One of the most innovative aspects of the Smartling TM system is that it is cloud-based, which appears to be the industry tendency for translation software (DePalma 2015; Toon 2015).

In terms of the object of research, as outlined in the previous section, for the purpose of this study, only translation problems related to or stemming from internationalization factors have been selected in our corpus of web apps. Additionally, since the source text is generally presented to translators in a TM system, certain challenges caused by it have also been taken into account. With the aim of illustrating these types of localization challenges, the following sections show the results from our corpus analysis.

4.2. Description of challenges

4.2.1. Challenges caused by software internationalization deficient practices

In terms of software internationalization practices, two principal challenges have been identified in the corpus. Firstly, challenges subsequent to not separating appropriately text and code. And secondly, challenges related to ambiguity.

4.2.1.1. User interface text separation

The extraction of text from code into a separate file (resource file) is an essential internationalization practice to avoid localization problems. When an application is not appropriately internationalized, it can happen that not all text is separated in a resource file. As a result, translators might not have access in the source text to certain localizable elements that are necessary for linguistic or stylistic correctness. This can be seen in the following example.

Plan changes will be applied to your next billing date on

The string above is segmented by the TM system as shown below.

TM SS: **Plan changes will be applied to your next billing date on.**

As explained in section 3.1, the red element behaves like a placeholder but it is not presented in the source text; therefore, translators might have difficulties in finding out what it is.

In this case, translators might assume that the proposition “on” is followed by a date with the format of month and day (and optionally, year), such as “July 1st, 2016”, but it could be followed instead by the segment “the last day of the month” or “Friday, July 1st”, to mention a few options. Without further context or access to isolated elements, translators might want to play it safe and avoid making any kind of grammar inconsistency and therefore, translate the strings as below:

TS: ***Los cambios de plan se aplicarán en la próxima fecha de facturación:***

4.2.1.2. Ambiguous strings

Context is fundamental for proper translation as explained above. Unfortunately, the internationalization process, though necessary, often causes the de-contextualization of user interface strings.

The two following segments are illustrative examples of de-contextualized strings that can pose quite a challenge to translators:

TM SS: **Logged in**
TM SS: **Removed the sign-in URL**

Without additional context, one would have serious difficulties translating these strings into Spanish (and probably into other languages) because there is not an explicit subject in any of them. In the first example, it is unknown whether one person has logged into a corporate email account or whether two people have logged into it. Similarly, in the second example, it is not clear whether “a person has removed the sign-in URL” as opposed to “a glitch has removed the sign-in URL”. Therefore, without further information, a wide variety of possible scenarios could apply depending on the web application features. As before, the safest and most neutral translation would probably be to provide an impersonal translation in Spanish:

TS: ***Sesión iniciada*** or ***Se inició sesión***
TS: ***URL de acceso eliminada*** or ***Se eliminó URL de acceso***

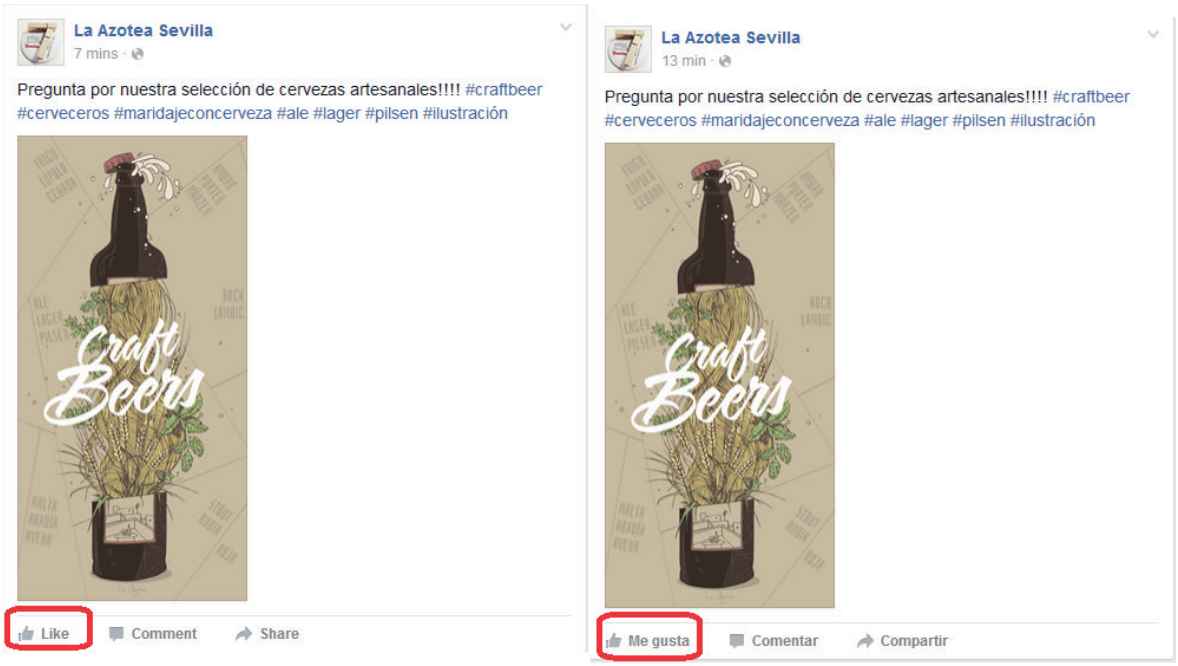
In this case, it was later found out that there was an implicit singular subject on the strings and therefore they could have been translated as "*Ha iniciado session*" and "*Ha eliminado la URL de acceso*" respectively. This challenge would have been easier to overcome had there been a comment or a description with additional information on the user interface strings, as explained in section 3.1.

Another ambiguous example found in the corpus is the following:

TM SS: **Like "name"**

As user interface strings are usually extracted without context in a separated file, possibly the first reaction of localizers of web applications would be to translate this string as the popular phrase of social media "Like" (in Spanish "*Me gusta*"), assuming that the web application they are translating has some kind of "Like" button. In social media sites or applications, the button or option "Like" is generally used to click or tap on a post (picture, comment, etc.) that a user likes on a given page. Facebook is the company that made this phrase popular and it has now spread to other social media websites like Pinterest and Instagram, although they use different icons (a thumb-up, a heart, etc.) to illustrate the "like" action:

Figure 3. Facebook Like button in English and Spanish



Alternatively, the string might simply refer to a prepositional phrase (preposition + noun), as in "Like Maria". In this case, unless translators are provided with further details, they might follow their intuition and risk translating the string as "*Me gusta*", potentially entering a translation mistake in the target text and consequently polluting the TM (which was one of the drawbacks of using TMs according to translators in LeBlanc's study). The string actually referred to "Like" as preposition (e.g. "*Como María*") and not to the social media phrase:

TS: ***Como*** **"name"**

In addition, the phrase "Like" introduces a wide variety of linguistic questions and concerns when translated into Spanish (and other languages). For instance, when people "like" a post on a given page, the text to describe that action shows the number of people that have liked that post, e.g. "5 Likes". This phrase would probably be translated into Spanish using "*Me gusta*" because that is the most widespread translation, but as the source phrase is pluralized ("Likes"), translators might have doubts on how to write it: 5 *Me gusta*, 5 *me gusta*, 5 "*Me gusta*", etc. On top of that, the string "likes" in isolation might also make translators wonder whether it is a verb (third person singular) as in the example above or a pluralized noun. And that would open a new set of challenges.

4.2.2. Challenges resulting from inadequate content internationalization

As mentioned in section 3.2, "writing for translation" (or content internationalization) refers to writing content with an international audience in mind, knowing that the written text is going to be sent for translation and that certain neutralization guidelines need to be followed. Apart from cultural and location-specific references, language structures that can constrain translation are also discouraged. In order to help developers write localizable content, certain technology companies, such as Mozilla or Microsoft, publish detailed internationalization and localization guidelines which stress the importance of avoiding assumptions with regards to language structure and singularities (Mozilla 1998; Microsoft 2016). Among the different guidelines, they remind developers that English word order and lack of gender might not replicate well (or at all) in other languages.

To examine this challenge, two examples from the corpus will be analyzed.

TM SS: **You found a %s!**

The placeholder "%s" could potentially be replaced by any singular noun which makes sense in terms of the application features (file, folder, error, match, etc.). Since there are two genders in Spanish (masculine and feminine), "a" could theoretically be

translated by "un" or "una" when there is no context or string description, as shown in this string:

TS 1: ***¡Has encontrado un/una %s!***

Or it could also be translated in other ways to avoid using "un/a".

TS 2: ***%s: elemento encontrado***
TS 3: ***¡Se encontró %s!***
TS 4: ***¡Has encontrado: %s!***

At application run-time, the placeholder would be replaced by an element (most likely depending on the user action), so the first segment would show in the user interface like *"¡Has encontrado un/una coincidencia!*" which certainly does not sound natural or correct in Spanish. The other segments would be displayed like "*coincidencia: elemento encontrado*", "*Se encontró coincidencia*" or "*¡Has encontrado: coincidencia!*", options clearly showing the user that this is an automatically generated text. Neither of them is ideal because the target user would probably feel the text as not written in their language. This problem could easily be solved if developers entered comments in the description field or if they used a different phrase or text to inform users that some element has been found in their system.

In the second case taken from the corpus, the translator will come across a similar challenge because writers have not taken into account the different constructions of languages:

TM SS 1: **for the last**

For this string, fortunately, there is some preview context. It is part of an application feature by which the user can generate reports on different measurable actions (saved items, posts published, items purchased, etc.) for a given period of time (e.g. week, 6 months, year, 2 years). The preview of the report feature would look something like this:

Show "saved items" for the last "week"

The underlined elements are four different translatable TM segments, and the quoted phrases are items on drop-down menus (that the user can choose from depending

on what kind of report they want to get). As a result, they are segmented in the TM system like this:

TM SS 1: **Show**
TM SS 2: **saved items (or posts published, purchases)**
TM SS 3: for the last
TM SS 4: **week (or 6 months, year, 2 years)**

The noun on TM SS 4 constrains the translation of "for the last" both by gender and by number in Spanish because the drop-down menu includes both masculine and feminine nouns (and plural masculine nouns) as options.

Considering that when an application is already developed and sent for translation, it is not likely that developers will want to make changes unless an important fix needs to be made, translators will have to work with the text sent for translation, and to creatively think for feasible solutions.

In the present example, a possible but not ideal solution would be to translate the string "for the last" as "*para anterior/es*". "Anterior" is a neutral adjective in Spanish working for masculine and feminine, and a good option to avoid too many alternative options (*para el/la último/a*) that would interfere with the user interaction with the application. It is advisable to use a precise and concise language when translating software to make sure that the use of the software is not hindered by language (Microsoft 2016).

4.2.3. Challenges related to the rigidity of segmentation in TMs

As seen on section 3.3, the process of segmentation can heavily constrain the localization process. As Austermühl (2006: 77) states:

Among the complaints from the translator community are the rigidity of source text structures, the dominance of the sentence or sub-sentence phrases as primary translation units, incompatibilities within one TM or between TM and term bases contents, faulty yet untouchable segments, the lack of creativity for the translator as autonomous text producer, the lack of co-text and context for the segments to be translated, and the lack of motivation or freedom to go beyond the simplistic source text structures and the preexisting translations imposed upon the translator by the TM system.

There are several challenges related to segmentation, but for the purpose of this study only two will be highlighted. Firstly, translators work with isolated units (segments) that potentially affect textual cohesion, especially because many TMs do not have a feature for segments merging which would be useful in many cases (as shown below) and because TMs generally do not allow for text structure changes. Secondly, and due to the singularities of every language, sometimes it seems unavoidable to

store segments incorrectly. Since sentences are split into segments, if source and target languages have different sentence structures, translators might have to play around with TM segments to produce a grammatically and stylistic correct sentence in the target language.

This two-fold segmentation problem can be exemplified with the following string found in the analyzed corpus:

Internal and external collaboration

This sentence is split into two segments in the TM system:

TM SS 1: **Internal and external**
TM SS 2: **collaboration**

Since merging segments is not allowed in the TM system used, it is almost impossible to store target segments correctly in the TM because Spanish usually requires that the noun precedes its adjective (its modifier) in a noun phrase. A potential translation would be:

TS 1*: **Colaboración interna***
TS 2*: **y externa***

As a result, the TM would save one translation unit as:

EN: **Internal and external – ES: *Colaboración interna***

And the second translation unit as:

EN: **collaboration – ES: *y externa***

Consequently, the TM would be polluted. There are other workarounds for this segment, but none of them would avoid this problem completely. Although this solution seems acceptable from a translation point of view, the pollution of the TM could affect productivity in future projects and hinder translation quality if they are not used cautiously (Bass 2006: 76).

5. Conclusions

Localization is everywhere. As a ubiquitous phenomenon, it is in the websites we read, in the devices we use and in the apps that make our life easier. However, translated content might not always sound natural in a given content, or might not even be correct, as exemplified earlier with the Zara translation mistake. In fact, some software manufacturers offer their users the possibility to report language problems they might come across (e.g. Google, Dropbox[6], Pinterest).

As this paper has tried to show, the localization of web applications can pose several challenges related to internationalization inadequate practices and to technological barriers (e.g. TM segmentation).

In terms of internationalization factors, two main sets of challenges have been identified: one related to the internationalization of the software itself (technical process), and another related to "content internationalization". The first set of challenges generally arises from the separation of text and code that, although essential for the localization process, can cause ambiguity problems (leading to "blind localization") and the mixing of code and text in the resource file if the application has not been correctly internationalized. The second set of challenges pointed out is inadequate content internationalization (or writing for translation), which can constrain the localization practice when writers do not consider cultural or language differences, e.g. structure.

As regards to technological barriers, TM system alteration of the text through segmentation can seriously limit translators and cause linguistic and stylistic errors as well as impact text cohesion. As mentioned earlier in this paper, localization is characterized by the intersection between translation and technology (Jiménez-Crespo 2010) so, given the outlook of this study, segmentation was taken into consideration as one of the technological challenges that localizers face.

Programmers, technical writers and localization engineers can help overcome these problems. If isolated UI strings (due to the separation of text and code) included comments or descriptions, the task of translators would be easier. Moreover, if writers wrote content with an international audience in mind and following internationalization guidelines, translators would not feel constrained by the source text. There are multiple guidelines and best practices available for content creators. However, the missing piece for implementing those in the content creation process might be found in businesses making certain decisions based on ROI considerations since localization is a process constrained by economic decisions. Similarly, if more resources were invested in making sure that TM systems help translators produce a correct and natural output instead of making them resort to workarounds, then text cohesion and time efficiencies would surely improve.

Both the fact that translators are only one of the actors involved in the translation production networks (Abdallah and Koskinen 2007) and that internationalization strategies often follow beyond the scope of the translation process serve as arguments for adopting a wider approach to the topics discussed in this paper. One could say that

translators at times lack the necessary freedom to create correct and natural target texts. As Harcz (2016: 23) argues:

> The extent to which liberty can be taken when rendering a script from one language to another is determined by several factors. Most notably: the targeted audience; the purpose (or intended use) of the translation; the subject matter of the document; the client who ordered the job; the specific instructions received when the order was placed; and my personal preferences and habits as a translator.

Furthermore, according to Alonso and Calvo (2015), in the future, tools will harmoniously complement translators' own capacities, as in the scenario depicted by these authors in their *trans-human translation* approach. Translators will interact with technology "as though this were really just an extension of their capabilities and creating a process with a social, creative and learning dimension" (Ibid.: 151), thus overcoming current tendencies towards the dehumanization of translation processes (Biau-Gil and Pym 2006) (cited in Alonso and Calvo 2015: 148).

Going back to Pickering's dialect of "resistance and accommodation" (in Olohan 2011: 344), one could argue that, although professional translators might be able to "accommodate" and work around challenges ("resistance") producing an acceptable output, these solutions might not be the most desirable options to successfully localize a product. This is particularly the case if one considers that localization entails "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" (LISA 1998: 3). Therefore, one might wonder if the challenges identified in this preliminary study prevent true localization from happening.

As a final remark, if localization quality evaluations considered the nature of source texts and the environments that translators work with when assessing quality, metrics might shed some light into what is causing certain translation errors. In the recent "Multidimensional Quality Metrics" project (QTLaunchPad 2014), it is advised that translation quality scorecards consider internationalization aspects:

> If Internationalization problems occur, they will negatively impact the quality score for the translated content, even though they are beyond the control of the translator. Any metrics used for translator assessment MUST exclude issue types that are beyond the control of the translator.

However, to consider internalization in localization quality evaluations, manufacturers and vendors should thoroughly audit source content and, unfortunately, this does not seem to be a standardized practice so far.

This paper has tried to flag the need to spend more resources in the internationalization of web applications, or to increase the "enterprise level of commitment to globalization strategies" (Dunne 2006: 6), considering that even some of the biggest

technology publishers have shortcomings on the internationalization process. The fact that few companies measure the ROI of localization and that they lack a comprehensive localization strategy (DePalma 2006) could explain why these shortcomings still persist. As Giammarresi argues: "Just as localization project pain is evidence of internationalization issues, internationalization issues are themselves symptoms of deeper, more systemic ills" (2011: 19).

Although localization professionals have been stressing for years the importance of thinking about localization during product development and of providing adequate tools for localization processes, according to the present study, translation problems still arise from deficient globalization processes. Hopefully, the day will come when most technology companies truly view localization as a strategic business decision that affects their revenue and their branding. When that happens, the challenges identified in this paper could be overcome.

6. Bibliography

- Abdallah, Kristiina and Koskinen, Kaisa (2007). Managing trust: Translating and the network economy. *Meta: Journal des Traducteurs / Meta: Translators' Journal* 52 (4), 673-687.
- Alonso, Elisa (2015). Analysing Translation Professionals in the Information Society and their Use and Perceptions of Wikipedia [online]. *JoSTrans: The Journal of Specialised Translation* 23 (January 2015), 89-116. <http://www.jostrans.org/issue23/art_alonso.php> [Accessed 25 March 2016].
- Alonso, Elisa (2016). Conflict, opacity and mistrust in the digital management of professional translation projects. *The International Journal of Translation and Interpreting Research* 8 (1), 19-29.
- Alonso, Elisa and Calvo, Elisa (2015). Developing a Blueprint for a Technology-mediated Approach to Translation Studies. *Meta: Journal des traducteurs Meta:/Translators' Journal* 60 (1), 135-157.
- Alonso, Elisa and De la Cova, Elena (2013). *Introducción a la localización. Materiales docentes* (An Introduction to Localisation. Training Materials). Seville (Spain): [self-edited].
- Amant, Kirk S. (2000). Expanding translation use to improve the quality of technical communication. *IEEE transactions on professional communication* 43(3), 323-326.
- — (2001). Cultures, computers, and communication: Evaluating models of international online production. *IEEE Transactions on Professional Communication* 44 (4), 291-295.
- — (2003). Designing effective writing-for-translation intranet sites. *IEEE Transactions on Professional Communication*, 46 (1), 55-62.

- — (2005). A prototype theory approach to Web site localization: an analytical method for technical communicators. *2005 IEEE International Professional Communication Conference Proceedings*, 101-107.
- Austermühl, Frank (2001). *Electronic tools for translators*. Manchester: St. Jerome.
- — (2006). Training translators to localize. In *Translation Technology and its Teaching*. Anthony Pym, Alexander Perekstenko and Bram Starink (eds.), 69-81. Tarragona: Intercultural Studies Group.
- Asnes, Adam (2007). Change your encoding, change your company. Getting started guide: Internationalization. *Multilingual* 18 (3), 5-6.
- Bass, Scott (2006). Quality in the real world. In *Perspectives on Localization*. Kieran J. Dunne (ed.), 69-84. Amsterdam-Philadelphia: John Benjamins.
- Bernal Merino, Miguel Ángel (2007). What's in a 'Game'?. *Localisation Focus The International Journal of Localisation* 6 (1), 29-38.
- Biau-Gil, José Ramón and Pym, Anthoy (2006). Technology and translation (a pedagogical overview). In *Translation Technology and its Teaching*. Anthony Pym, Alexander Perekstenko, and Bram Starink (eds.), 5-19. Tarragona: Intercultural Studies Group.
- Bowker, Linda and Pearson, Jennifer (2002). *Working with specialized language: a practical guide to using corpora*. London-New York: Routledge.
- Bowker, Linda (2005). Productivity vs quality. A pilot study on the impact of translation memory systems. *Localisation Focus* 4, 13-20.
- Brown, M. Katherine (2003). Trends in Writing for Translation. *Multilingual* 14 (7), 4-8.
- Bundgaard, Kristine, Christensen, Tina Paulsen and Schjoldager, Anne (2016). Translator-Computer Interaction in Action—an Observational Process Study of Computer-Aided Translation. *Journal of Specialised Translation* 25, 106-130
- Castells, Manuel (2005). Internet y la sociedad red. In *Por otra comunicación: los media, globalización cultural y poder*. Denis de Moraes (ed). 203-228. Barcelona: Intermón Oxfam Editorial.
- Christensen, Tina Paulsen and Schjoldager, Anne (2011). The Impact of Translation-Memory (TM) Technology. In *Human-Machine Interaction in Translation: Proceedings of the 8th International NLPCS Workshop.* Bernadette Sharp, Michael Zock, Michael Carl and Arnt Lykke Jakobsen (eds). 119-130. Copenhagen: Samfundslitteratur.
- Colominas, Carme (2008). Towards chunk-based translation memories. *Babel* 54 (4), 343-354.
- Corpas Pastor, Gloria and Seghiri Domínguez, Miriam (2006). *El concepto de representatividad en la Lingüística del Corpus: aproximaciones teóricas y metodológicas*. Documento técnico BFF2003-04616 MCYT/TI-DT-2006-1. Universidad de Málaga.
- Corte, Noelia (2002). Localización e internacionalización de sitios web. *Tradumàtica* 1, 1-8.

- Cronin, M. (2003). *Translation and globalization*. London-New York: Routledge.
- De la Cova, Elena (2014). Localización de sitios web: ¿a quién van dirigidos? *IV Congreso de la Sociedad Española de Lenguas Modernas*. Sevilla: Sociedad Española de Lenguas Modernas
- — (in progress). *Localization problems of Online Help* (temporary title). Doctoral Thesis, Universidad de Sevilla
- DePalma, Donald (2006). Quantifying the return on localization investment. In *Perspectives on Localization*. Kieran J. Dunne (ed.), 15-36. Amsterdam-Philadelphia: John Benjamins.
- — (2015). The cloud: a translation business essential. *Multilingual* 26 (1), 26-27
- Diéguez Morales, Mª Isabel y Lazo Rodríguez, Rosa Mª (2011). El español en Internet: aciertos y errores en sitios web localizados del inglés. *Onomázein: Revista de lingüística, filología y traducción de la Pontificia Universidad Católica de Chile* 24, 299-326.
- Dragsted, Barbara (2006). Computer-aided translation as a distributed cognitive task. *Pragmatics & Cognition* 14 (2), 443-464.
- Dunne, Keiran (2015). Localization. In *Routledge Encyclopedia of Translation Technology*. Sin-wai Chan (ed.), 550-562. London-New York: Routledge.
- — (2006). A Copernican revolution. In *Perspectives on Localization*. Kieran J. Dunne (ed.), 1-11. Amsterdam-Philadelphia: John Benjamins.
- Edwards, Kate (2015). The death of context. *Multilingual* 26 (8), 20-21.
- Esselink, Bert (2000). *A practical guide to localization*. Amsterdam-Philadelphia: John Benjamins.
- — and Cadieux, Pierre (2002). GILT: Globalization, Internationalization, Localization, Translation. *The LISA Newsletter: Globalization Insider XI* (1.5).
- — (2003). The evolution of localization. The Guide to Localization. *Multilingual* 14 (5), 4-7.
- Fenstermacher, Hans (2006). Authors, localizers and language barriers. *Multilingual* 17 (1), 82
- Flint, Patricia, Van Slyke, Melanie Lord, Stärke-Meyerring, Doreen and Thompson, Aimee (1999). Going online: Helping technical communicators help translators. *Technical communication* 46 (2), 238-248.
- Giammarresi, Salvatore (2001). Strategic views on localization project management. In *Translation and localization project management: the art of the possible*. Kieran J. Dunne, and Elena S. Dunne (eds.), 17-51. Amsterdam-Philadelphia: John Benjamins.
- Harcz, Daniel B. (2016). Taking liberty as a translator. *Multilingual* 27 (1), 23.
- Herrmann, Achim and Sachse, Florian (2005). Internacionalización de aplicaciones de software. In *Traducción y localización. Mercado, gestión y tecnologías*. Detlef Reineke (ed.), 45-70. La Palmas de Gran Canaria: Anroart.
- Hoft, Nancy L. (1995). *International technical communication: How to export information about high techno*logy. New York: John Wiley & Sons, Inc.

- Horvath, Alan (2007). Think internationalization in everyday design. Getting started guide: Internationalization. *Multilingual* 18 (3), 3-4.
- Hurtado Albir, Amparo (2001). *Traducción y traductología*. Madrid: Cátedra.
- Jiménez Crespo, Miguel Ángel (2008). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. Tesis doctoral, Universidad de Granada
- — (2009a). The effect of Translation Memory tools in translated Web texts: Evidence from a comparative product-based study. *Linguistica Antverpiensia* 8, 213–234.
- — (2009b). El uso de corpus textuales en localización. *Tradumàtica: traducció i tecnologies de la informació i la comunicación* 7, 1-15
- — (2010). The intersection of localization and translation A corpus study of Spanish original and localized web forms. *Translation and Interpreting Studies* 5(2), 186-207.
- — (2011a). To adapt or not to adapt in web localization: a contrastive genre-based study of original and localized legal sections in corporate websites [online]. *Jostrans: The Journal of Specialised Translation* 15, 2-27. <http://www.jostrans.org/issue15/art_jimenez.php> [Accessed 15 March 2016].
- — (2011b). A corpus-based error typology: towards a more objective approach to measuring quality in localization. *Perspectives, Studies in Translatology*, 19 (4): 315-338.
- — (2012). Loss or Lost in Localization: A Corpus-Based Study of Original and Localized Non-Profit Websites [online]. *Jostrans: The Journal of Specialised Translation* 17, 136-165. < http://www.jostrans.org/issue17/art_jimenez.php> [Accessed 16 March 2016].
- — (2013). *Translation and web localization*. London-New York: Routledge.
- Kamerer, Laurie (2003). Avoiding a US-centric Writing Style. *Multilingual* 14 (7), 22-23
- Keller, Nicole (2011). Creating translation-oriented source documents, *Multilingual* 22(7), 43.
- Kiraly, Don C. (1995). *Pathways to translation: Pedagogy and process* (No. 3). Ohio: Kent State University Press.
- Lagoudaki, Elina (2006). Translation memories survey 2006: Users' perceptions around tm use. *Proceedings of the ASLIB International Conference Translating & the Computer* (28) 1, 1-29.
- LeBlanc, Matthieu (2013). Translators on translation memory (TM). Results of an ethnographic study in three translation services and agencies. *Translation and Interpreting* 5 (2), 1–13.
- LISA (1998). *LEIT phase I final report. LISA Education Initiative Taskforce (LEIT),* Report November, 22.
- Lommel, Arle and Ray, Rebecca (2007). *LISA globalization industry primer*. Romainmótier*: LISA*.

- Lörscher, W. (1991). Translation performance, translation process, and translation strategies: A psycholinguistic investigation. *Target* 4 (1), 126-129.
- Ludwigsen, Scott, Williams, Jeff, y Polis, Jennifer (2009). The Guide to Translation & Localization. Portland: lingo-systems.
- Massion, Francois (2010). Automating translation quality management. *Multilingual,* 21 (1), 41.
- Mata, Manuel (2005). Localización y traducción de contenido web. In *Traducción y localización. Mercado, gestión y tecnologías.* Detlef Reineke (ed.), 187-252. Las Palmas: Anroart.
- — (2009). Algunas pautas para el tratamiento de imágenes y contenido gráfico en proyectos de localización (II). *EntreCulturas* 1, 533-569.
- Medina, Ana and Ramírez, Cristina (2015). La localización de la sección" productos" en sitios web de empresas exportadoras agroalimentarias [online]. *InTRAlinea: Online Translation Journal*. <http://www.intralinea.org/specials/article/la_localizacion_de_la_seccion_productos_en_sitios_web_de_empresas> [Accessed 2 April 2016]
- McEnery, Tony and Wilson, Andrew (2001). *Corpus linguistics: An introduction*. Edinburgh: Edinburgh University Press.
- McKethan Kenneth and White, Graciella (2005). Demystifying software globalization[online]. *Translation Journal* 9 (2), 1-8. <http://translationjournal.net/journal/32global.htm> [Accessed 2 April 2016]
- Meersseman, Yann. The customer makes the difference. In *LISA Best Practice Guide. Quality Assurance- the client perspective.* LISA (ed.), 35-37. Switzerland: The Localization Industry Standard Association.
- Mossop, Brian (2006). Has computerization changed translation?. *Meta: Journal des traducteursMeta:/Translators' Journal* 51 (4), 787-805.
- Nord, Christiane (2005). *Text analysis in translation: Theory, methodology, and didactic application of a model for translation-oriented text analysis*. Amsterdam-New York: Rodopi.
- Oliver, Antoni, Moré, Joaquim and Climent, Salvador (2008). *Traducción y tecnologías*. *Barcelona : Editorial UOC, 2008.*
- O'Hagan, Minako (2009). Evolution of user-generated translation: Fansubs, translation hacking and crowdsourcing. *Journal of Internationalisation and Localisation* 1 (1), 94-121.
- Olohan, Maeve (2004). *Introducing corpora in translation studies*. London-New York: Routledge.
- — (2011). Translators and translation technology: The dance of agency. *Translation Studies* 4 (3), 345–357.
- PACTE Group (2011). Results of the validation of the PACTE Translation Competence model: Translation project and dynamic translation index. En *Cognitive Explorations of Translation*. Sharon O'Brien (ed.), 30-53. London-Nueva York: Continuum Studies in Translation.

- Presas, M. Lluïsa (1996). *Problemes de traducció i competència traductora. Bases per a una pedgagotia de la traducción.* Tesis doctoral, Universitat Autònoma de Barcelona.
- Pym, Anthony (2004). *The Moving Text. Localization, Translation, and Distribution*. Amsterdam and Philadelphia: John Benjamins.
- — (2010). *Exploring Translation Theories*. London-New York: Routledge.
- — (2011a). Website localization. In *The Oxford Handbook of Translation Studies*. Kirsten Malmkjaer and Kevin Windle (eds), 410-23. Oxford: Oxford University Press.
- — (2011b). What technology does to translating. *Translation and Interpreting* 3 (1), 1–9.
- Reppen, Randi (2010). What are the key considerations?. In *The Routledge Handbook of Corpus Linguistics.* Anne O'Keeffe and Michael McCarthy (eds.), 31-37. London-New York: Routledge.
- Rojo, Ana (2013). *Diseños y métodos de investigación en traducción*. Madrid: Editorial Síntesis.
- Roturier, Johann, and Lehmann, Sabine (2009). How to treat GUI options in IT technical texts for authoring and machine translation. *The Journal of Internationalization and Localization* 1(1), 40-59.
- Safar, Libor and Machala, Jiri (2010). Best practices in localization testing. *Multilingual 21(1)*, 28-32.
- Sandrini, Peter (2005). Website localization and translation. In *EU-High-Level Scientific Conference Series: Challenges of Multidimensional Translation* . Heidrun Gerzymisch-Arbogast and Sandra Nauert (eds.), 131-138. Saarbrücken: MuTra Conference Proceedings.
- — (2008). Localization and Translation. In LSP Translation Scenarios. Heidrun Gerzymisch-Arbogast, Gerhard Budin and Gertrud Hofer (eds.), 167. Wiesbaden: ATRC Group.
- Schäler, Reinhard (2007). Reverse localisation. *Localisation Focus: the International Journal of Localisation* 6 (1), 39-48.
- — (2008). Linguistic resources and localization. In *Topics in Language Resources for Translation and Localization*. Elia Yuste (ed.), 195-225. Amsterdam-Philadelphia: John Benjamins.
- Sinclair, John (2002) Preface. In *Small Corpus Studies and ELT. Theory and Practice*. Mohsen Ghadessy, Alex Henry and Robert L. Roseberry (eds.), xii-xv. Amsterdam-Philadelphia: John Benjamins.
- Sikes, Richard (2010). Five types of localization testing. *Multilingual* 14 (7), 4-8.
- Sichel, Barb (2009). Planning and writing for translation. Getting Started Guide: Writing for translation. *MultiLingual*, 20 (7), 3-4
- Somers, Harold (2003). *The Translator's Workstation Computer and Translation: A Translator's Guide*. Amsterdam-Philadelphia: John Benjamins.
- Toon, Alison (2015). Going global via the cloud. *Multilingual*26 (1), 26-27

- Toury, Gideon (2010). What's the problem with 'translation problem'? *Meaning in Translation*. Barbara Lewandowska-Tomaszczyk and Marcel Thelen (eds.), 231-252. Frankfurt am Main: Peter Lang.
- Vine, Andrea S. (2002). Internationalization in Software Architecture and Design. *Multilingual* 13 (3), 15-22
- Wilss, Wolfram (1982). *The science of translation: Problems and methods*. Tübingen : Gunter Narr Verlag
- Wu, Ye and Offutt, Jeff (2002). Modeling and testing web-based applications. *GMU ISE Technical ISE-TR-02-08*. Technical reports, George Mason University.
- Yang, Yanxia (2007). Extending the user experience to localized products. In *Usability and Internationalization. Global and Local User Interfaces*. Nuray Aykin (ed.), 285-292. Springer Berlin Heidelberg.
- Online Resources
- EAGLES. *What is a corpus?* EAGLES Text Corpora Working Group Reading Guide <http://www.ilc.cnr.it/EAGLES96/corpintr/node13.html#SECTION00040000000000000000> [Accessed 2 February 2016]
- GALA. Terminology. <http://www.galaglobal.org/> [Accessed 2013]
- GALA. About GALA. <http://www.galaglobal.org/> [Accessed 5 March 2016]
- La Vanguardia. *Zara causa polémica en Alemania por la venta de unas 'sandalias de esclava'*. <http://www.lavanguardia.com/de-moda/moda/mujer/20150409/54429772886/zara-polemica-alemania-sandalias-esclavas.html> [Accessed 25 February 2016]
- Lommel, Arle, Burchardt, Aljoscha, Uszkoreit, Hans y Melby Alan (2015). Multidimensional Quality Metrics (MQM) Issue Types. <http://www.qt21.eu/mqm-definition/issues-list-2015-12-30.html#internationalization> [Accessed 10 April 2016]
- — (2016). Multidimensional Quality Metrics (MQM) Definition. <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html#issue_types> [Accessed 16 October 2016]
- Microsoft Developer Network. *Coding for Internationalization*. <https://msdn.microsoft.com/es-es/library/ms903791.aspx> [Accessed 10 April 2016]
- Mozilla. Mozilla Localizability Guidelines. <http://www-archive.mozilla.org/docs/refList/i18n/l12yGuidelines.html> [Accessed 10 April 2016]
- QTLaunchPad (2014). High-Level Structure. <http://www.qt21.eu/launchpad/node/1332> [Accessed 15 April 2016]
- Roberts, Adam (2014). Span (HTML element). <http://www.sitepoint.com/web-foundations/span-html-element/> [Accessed 15 April 2016]
- Twitter. #sklavensandalen page <https://twitter.com/hashtag/sklavensandalen> [Consulta: 15 marzo 2016]
- W3C. <https://www.w3.org/International/> [Accessed 15 April 2016]
- — Your first look at JavaScript. <https://www.w3.org/wiki/Your_first_look_at_JavaScript> [Accessed 10 April 2016]

Notes

1. LISA was a non-profit association founded in 1990 aimed at "promoting the localization and internationalization industry and providing a mechanism and services to enable companies to exchange and share information on the development of processes, tools, technologies and business models connected with localization, internationalization and related topics" (Lommel and Ray, 2007, p. 53)

2. The Globalization and Localization Association (GALA) is a leading trade association for the language industry that supports "member companies and the language sector by creating communities, championing standards, sharing knowledge, and advancing technology." (GALA, 2016)

3. "The W3C Internationalization (I18n) Activity works with W3C working groups and liaises with other organizations to make it possible to use Web technologies with different languages, scripts, and cultures" (W3C 2014).

4. "JavaScript is a text-based language that does not need any conversion before being executed. (…) It is executed instantly by a type of program that interprets the code called a parser." (W3C, 2012)

5. "Smartling provides an innovative, cloud-based translation management platform that lets companies quickly translate and deploy their websites and dynamic web applications, mobile applications and business documents across the entire corporate digital infrastructure" (Multilingual, 2015)

6. Dropbox Help Request: https://www.dropbox.com/support/s/215/3281254/c/214