

To adapt or not to adapt in web localization: a contrastive genre-based study of original and localised legal sections in corporate websites

Miguel A. Jiménez-Crespo, Rutgers University, The State University of New Jersey, USA.

ABSTRACT

Since the early 90's, the localisation industry has striven to produce non-culture-specific texts that can be easily localised into most languages. Nevertheless, international websites include sections, such as legal disclaimers or privacy policies, that preferably need to be adapted in order to be fully effective and increase the credibility of the website (Kenny and Jones 2007). This study explores these two seemingly contradictory perspectives through a comparable corpus analysis of original and localised legal sections in corporate websites. Following a genre-based approach (Swales 1990; Bhatia 1993; Gamero 2001), the main analysis concentrates on macrostructural differences and representative conventional linguistic forms associated with rhetorical moves. The analysis shows significant differences in the prototypical macrostructures of original and localised texts, as well as an impact on their terminology and phraseology. As far as the adaptation is concerned, only 32.60% of websites were somewhat adapted to the Spanish target legal system, while the rest were localised but not legally adapted. The results shed some light on the question of whether current industry strategies favor single internationalised vs. adapted localisations and on the inevitable effect of source text structures and phraseology on the final localised website.

KEYWORDS

Web localisation, localization, translation, adaptation, corpus-based translation studies, web genres, genre theory

1. Introduction

For most non-English speaking cultures, the ever-increasing digital world cannot be understood without the mediation of web localisation processes. Millions of web users interact daily with localised web content and browsers.¹ In fact, it could be claimed that localised webs might represent the most-used translated texts globally. From a translation perspective, the features of translated texts are, and have been, widely studied since the emergence of corpus-based translation studies (e. g. Baker 1993, 1995; Laviosa 1998), and nowadays, localised texts should be brought to the forefront of this discussion. So far, this translation-mediated communication process has not yet been granted the attention it deserves from both a theoretical and an empirical perspective (Pym 2003, 2005, 2009; Jiménez Crespo 2009b, 2010; Dunne 2006). Among the many issues that demand more detailed analysis, this paper focuses on the special *adaptation* component that the industry claims is the main difference between localisation and translation. The focus of the analysis is legal texts embedded in websites, a potential probe into whether web

texts are localised maintaining macro and microstructures of source texts (Jiménez-Crespo 2009b), or fully adapted to the target sociocultural and legal context. The adaptation of legal sections in localisation to the target sociocultural contexts is of paramount importance as it can impact the end user's confidence in the entire site (Kenny and Jones, 2007). As such, the mere translation of the legal content without appropriate adaptation could be detrimental to the goal of the target text in the sociocultural context of reception. Ideally, this adaptation should be done in consultation with legal experts, as they are also responsible for the drafting of source legal sections in websites (Garrand 2001).

The localisation industry, in its attempts at defining localisation as a process that goes beyond translation, normally claims the existence of this *adaptation* component to fulfill the expectations of local users (e.g. Esselink 2000; Önorm D 1200; Microsoft 2003; LISA 2003, 2007:14; Dunne 2006:4; Industrie Canada, 1999:48; Sighn and Pereira 2005). This cultural and technical adaptation is widely seen and presented as the most important differential factor between translation and localisation. However, since Nida's proposed dynamic and semantic equivalence models as well as the emergence of functionalist approaches (e.g. Reiss and Vermeer 1984; Nord 1997), translation adaptation to the receiver's context or expectations is regarded as inherent in all target-oriented translation processes.

Methodologically, this issue is researched through a contrastive genre-based analysis of localised (LT) and original legal texts (OT) in corporate websites. These texts are understood as a conventionalised *move* (Swales 1991) or *communicative section* (Gamero 2001) in the corporate website digital genre. The study analyses contrastively the average frequencies of the many textual sections and subsections in these Spanish OT and LT into this same language, such as privacy policies, legal disclaimers, terms and conditions, etc. Additionally, and given that in legal texts there is a high level of conventionalisation in the phraseological units associated to the different *steps* or *moves* (Borja Albí 2000, 2005), a contrastive phraseological analysis is performed in a second stage.

2. Conceptualising web localisation

Among all branches of localisation, web localisation has without any doubt the largest volume of translation (LISA 2007). It can be defined as a complex communicative, textual, cognitive and technological process by which interactive multimedia web texts are modified in order to be used by a target audience whose language and sociocultural context are different from those of original production (Jiménez-Crespo 2010). Web localisation developed by modeling and adapting certain processes and practices already established in software localisation (Dunne 2006; Yunker 2003: 30), but due to the explosion in the volume of information shared

through the WWW, the economic impact of the former is currently far greater than of the latter (Schäler 2005). Its relative youth led several scholars to coin different terms during the last decade, such as *e-localization* (Cronin 2003), *web globalization* (Yunker 2003), *content localization* (Esselink 2006) or *web-content localization* (Mata Pastor 2005). Nevertheless, a review of recent literature clearly shows that the most conventional term used by both translators and practitioners alike is *web localization*, and given the need for a common and stable metalanguage of translation and localisation (Chesterman 2005; Mazur 2008), this will be the term used henceforth. Furthermore, and in order to clarify any conceptual ambiguities, it should be mentioned that web localisation concentrates exclusively on multimedia texts stored and distributed through the WWW, but it does not include texts from other Internet-mediated communicative exchanges, such as chats, SMS or forums (O'Hagan and Ashworth 2003).

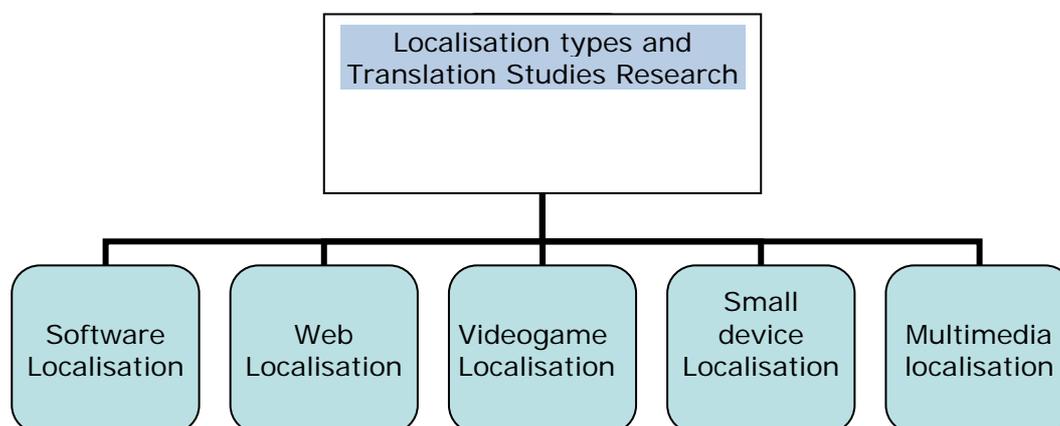


Figure 1. Different areas of research in Localisation Studies.

Most localisation processes, as shown in Figure 1, share several characteristics, such as the digital nature of the text, the presentation on screen, the interactive nature of texts or the necessary collaboration with localisation engineers and developers to produce the final target product. Nevertheless, there are stark differences in the way the actual textual segments are stored, the programming or markup languages used or the potential variation in textual types and genres (Jiménez-Crespo 2008b, 2009b). As an example, most software products entail a relatively standardised textual genre (Aüstermuhl 2007), videogame localisation also deals with a limited number of genres (Mangiron and O'Hagan 2006), but on the contrary, most web digital genres are complex genres (Martin 1995; Hanks 1986), that is, genres that can potentially incorporate a wide range of secondary genres, such as online purchase contracts in e-commerce websites. This is what Bhatia (1986) or Martin (1995:25) referred to as *genre embedding*. Consequently, despite the fact that most widely used digital genres are nowadays highly conventionalised (Shepherd and Waters 1998; Shepherd *et al.* 2005; Santini 2007; Jiménez

Crespo 2008b), a web localiser can potentially encounter a huge variety of secondary genres embedded in any website.

It is also clear that this new process needs to be contextualised in its relation to the Internet and the WWW. The latter has not only led to the emergence of this new modality, but it has also revolutionised translation and business practices around the world (LISA 2007; Gouadec 2007). It should be mentioned that not all texts distributed on the WWW are the result of the new textual and communicative model that emerged through the hypertextual revolution (Storrer 2002; Crystal 2001). The WWW allows any text created or converted into digital format to be distributed through the WWW. As an example, an instruction booklet for any product can be uploaded in a website without modifications; a governmental website normally offers official scanned documents in html or pdf format. These types of texts are what Angelika Storrer (2002) refers to as *e-texts*: sequentially organised printed documents that are simply uploaded and made available on the WWW. These e-texts can also be conceptualised as digital secondary genres (Martin 1995), as they can be randomly embedded in any hypertext. As such, processing these documents cannot be per se the object of study of web localisation, but rather, the overall digital genre structure that allows for this genre embedding, that is, the corporate or social networking site as a whole. Additionally, Storrer considers *hypertexts* as the new textual and communicative model that appears exclusively on the WWW.² They can be defined as networks of textual nodes and links that serve a distinct textual function and address a comprehensive, global topic. These hypertexts are open, as the developer can add any other nodes or textual segments at any time. In hypertext theory, *nodes* are defined as subunits that form independent unitary communicative chunks, such as textual segments, navigation menus, graphics, pictures, ad banners, flash files, etc. (Codina 2003).³ Thus, this paper proposes that hypertexts can be defined as the prototypical object of study of web localisation following what Toury (1995) and Holmes (1988) would consider a *restricted* theoretical area inside T&S.

Moreover, due to storage, retrieval and screen presentation purposes, each webpage in a hypertext is in turn subdivided into *interface text* and *content text* (Prince and Price 2002). The former includes all textual segments whose function is to help users navigate the hypertextual structure. As such, these types of texts are repeated throughout the website and they help negotiate the global coherence in a complete website (Fritz 1998; Storrer 2002). These textual segments include navigation menus, search functions or web page descriptions and content tags in the headings <head> </head>. Interface texts tend to be more conventionalised as digital genres are gradually being highly conventionalised with a common structure (Santini 2007; Nielsen and Loranger 2007; Jiménez-Crespo 2008b). On the other hand, *content text* can be defined as the unique differentiated textual content that makes each web page a storage unit as summarised in the webpage title. As an

example, in any conventional *contact us* page, the contact information for the party responsible for the website can be defined as the content text, while the rest of the text, such as navigation menus or banner ads is the interface text. As an example, digital newspapers constitute a new digital genre that evolved parallel to the expansion of the WWW (Shepherd and Watters 1998). Nevertheless, any piece of printed news simply posted in a digital paper could not be defined as a textual exemplar that is exclusively dependent on the medium; its translation process would be similar to the translation of any other printed piece of news.⁴

As far as the localisation process, and from a Translation Studies perspective, web localisation can be defined mostly as an *instrumental* (Nord 1997) or *covert* (House 1997:111) process in which the goal is for end users to interact with the translated text as if the text was directly produced in the target language. This is implicitly indicated in the goals for localisation laid out by the Localization Industry Standards Association (2004, 2003), as websites are to be received as “locally made products” or look like they have been developed in-country. In this translation type, end users are unaware that they are in fact interacting with a translated text, and the adaptation to the cultural and linguistic expectations of the target user is of utmost importance. Nevertheless, the legal texts under study represent a completely different translation type, as legal translation requires a *documentary* (Nord 1997) or *overt* (House 1997) translation type. This means that the translation is presented as such and normally, the faithfulness to the source text becomes an essential aspect. Therefore, while translators have a wider range of possibilities while adapting the website to the expectations of the target audience, they face a completely different translation process in these legal sections. This documentary nature is sometimes implicitly formulated in legal texts, normally indicating that the English source version is the only valid one in the case of legal disputes. This poses an interesting challenge to localisers as they need to handle different translation types during the course of a web localisation project. The results from this study will help answer the question about whether or to what extent these sections are in fact translated differently or not.

Now that web localisation has been defined and contextualised in the realm of Translation Studies, the next section reviews legal texts in websites from a textual genre perspective in order to clarify the methodological approach taken.

3. Legal information in websites: a genre description

Genre-based approaches to legal translation have been extremely productive during the last decade. This is due to the fact that legal genres are highly structured and conventionalised (Alcaraz and Hughes 2002; Borja Albí 2005; Cao 2007). Contrastive genre-based research of legal genres have been extremely beneficial to translation trainers, practitioners

and researchers as it allows them to analyze and adjust not only the macrostructure of the source text to the conventionalised macrostructure⁵ of the same legal genre in the target sociocultural context, but also the phraseological and terminological conventions associated to any of the many moves, steps or textual blocks. This high level of conventionalisation of legal structures in websites can be witnessed by the existence of standardised privacy or terms and conditions in published books (i.e. Gonzalez et. al 2004; American Bar Association 2007) or online interactive generators that can be directly used in any website.⁶ Recent research following this contrastive genre-based approach has led to the development of corpora with the most translated legal genres, such as the GITRAD corpus (Borja Albi 2007), a first step towards the description and analysis of the prototypical macrostructures and microstructures of these genres in several languages and sociocultural contexts.

Methodologically, these contrastive studies follow an analysis continuum starting from the superstructure and macrostructure (Göpferich 1995), usually describing and then contrasting the prototypical genre's macrostructure. In order to research these prototypical textual structures, any given genre is subdivided into recurring *sections>moves>steps>substeps* (Swales 1990), *triad>keys* (Paltridge 1997) or *communicative blocks>communicative sections>significant units>significant subunits* (Gamero 2001). The frequencies of each textual section identified by the researcher are recorded in order to identify their level of conventionalisation. In a later stage, a microstructural analysis can also be performed in which conventional linguistic forms⁷ that recurrently appear in each macrostructural section are identified and contrasted between both cultures.

Following this approach, the adaptation claim by the localisation industry will be researched through a contrastive study of the prototypical macrostructures in Spanish original and localised web texts. This will be followed by a microstructural contrastive analysis which focuses on conventionalised phraseology that appears in a representative selection of textual sections.

4. Empirical Study: Methodology

The comparable corpus of Spanish original (OT) and localised legal texts (LT) was extracted from the Comparable Web Spanish corpus compiled by Jiménez-Crespo (2008a). This wider corpus included 95 localised websites for Spain from the largest US companies, as well as a representative collection of 175 original corporate websites from Spain. It was collected in November of 2006. The subcorpus under study includes all pages in this larger comparable corpus with legal content, such as legal disclaimers, privacy policies, terms and conditions and copyright-trademark pages. The Spanish original section of the corpus under study comprises 64 legal web

pages, with 57,718 words and 4776 different tokens. The localised section has 65 legal web pages, with 112,319 words and 7495 different tokens. The number of web pages is higher than the number of total websites given that many websites have two or more pages for legal content, such as a page for a privacy policy and a page for terms and conditions.

	Original web legal corpus	Localised web legal corpus
Words	57,718	112,319
Types	4776	7495
Webpages	64	65
Websites	54	47

Table 1. Description of the comparable web legal corpus of corporate websites.

In order to contrastively analyze the macrostructure of these two textual subcorpora, all pages were analyzed as a single *legal* move in each website. This was necessary given that the distribution of content is normally uneven among these pages. That is, a privacy policy page might include information about the terms and conditions, and a legal disclaimer page might include all other legal information regarding privacy and terms, etc. The different moves, steps and substeps were carefully analyzed and the frequency of appearance of each of them was recorded. This means that the frequencies recorded indicate the appearance per site for each constituent textual section identified.

5. Results and discussion

The first analysis, in Table 1, shows that the average number of words per site and per page, and the values of both are much higher in LT. The average number of words per web page with legal content in OT is 901.84, while LT shows an average of 1727.98 words, almost double the value in original ones. This finding points out that LT are on average much longer, and consequently, their macrostructure will inevitably show a higher volume of constituent moves and steps. In order to situate this result in the context of the global website, the average number of words per page in the overall Spanish Web Comparable Corpus is 258.87 in the original section and 416.07 in the localised one (Jiménez-Crespo 2008a: 273). Thus, for both sections, legal web pages normally contain almost four times more words than the rest of webpages in corporate sites. Furthermore, if the number of words in all legal texts are compounded per website, localised legal sections show an average of 2415.69 words, while original sites with legal content show on average 1074.94 words per site [+224.72% difference]. The longer formulation in localised websites would in principle lessen their usability and readability as style guidelines and empirical usability research recommend briefness and conciseness in web pages (e.g. Nielsen and Loranger 2006; Jeney 2007; Price and Price 2002). Moreover, according to usability research it is recommended to

avoid page scrolling, because users normally avoid this process and move to other pages (Nielsen 1999; Price and Price 2002: 147).

Legal comparable subcorpus	Word average per web page in legal subcorpus	Word average per web page in Spanish Web Comparable Corpus	Average per website with legal moves-steps
Original	901.84	258.87	1068.85
Localised	1727.98	416.07	2389.76

Table 2. Number of words per page with legal content and per corporate website.

A potential explanation of this difference might be due to the different legal content in the source and target sociocultural contexts. In the USA, web privacy issues are self regulated by companies themselves under the guidance by the Federal Trade Commission (Liu and Arnett 2000), while in the Spanish Legal system, web privacy is regulated by the 1999 Spanish Data Protection Act. This means that US websites are required to explicitly formulate a full privacy policy block, while Spanish sites only have to indicate that their practices are in compliance with the applicable Spanish law. This again indicates that to some extent, the texts are not adapted as the macrostructure from the source text is maintained.

This first analysis has shown that the length of localised texts tends to be on average twice that of original ones. With this in mind, the next section explores exactly which moves and steps might be contrastively under-or over-represented in both corpora.

5.1. Contrastive macrostructural analysis

For the next analysis, all OT and LT were manually examined and the potential constituent moves and steps were identified. After this descriptive analysis of all potential legal moves and steps, the macrostructure of each legal text was examined, each previously identified move or step was tagged and its frequency was recorded. Table 3 shows the contrastive analysis of the frequency of all moves, steps and substeps in these legal texts. Following previous studies in this area (Jiménez Crespo 2008a, 2008b; Nielsen and Tahir 2002), three main moves were identified: legal disclaimers (M1), privacy policy (M2) and terms and conditions (M3). In each move, all different steps and substeps were identified and recorded. As an example, in the *legal disclaimer* move (M1), ten different steps were identified, such as *introduction* (S1-1), *acceptance of legal terms* (S1-2), *company registration* (S1-3). Globally, ten steps were identified in M1, eleven steps in *privacy policy* (M2) and thirteen in *terms and conditions* (M3). For many of the steps, several substeps were also identified, and these were marked with a consecutive

letter of the alphabet. As an illustration, the first step in the *legal disclaimer* move includes two substeps, a welcoming statement to the legal webpage (S1-1-a) and an appeal to read the legal text in its entirety (S1-1-b). Only eighteen substeps were recorded in Table 3, but nevertheless, it should be mentioned that the identification of these secondary textual blocks should be understood as an open framework in which more substeps could be potentially included.

The frequencies recorded in Table 3 are indicative of the appearance in the web pages collected in the legal corpus. In the wider Spanish Web Comparable corpus, as previously reported by Jiménez-Crespo (2009b), the three basic legal sections show much higher frequencies in localised corporate websites than in original Spanish ones: *privacy policies* appear in 70.52% of localised corporate websites and in 13.37% of original ones, *terms and conditions* in 38.94% of LT and in 4.65% of OT and *legal disclaimers* in 47.36% of LT and in 27.90% of OT. This finding is also consistent with the results from another study (Robbin and Stylianou 2003) that concluded that the most consistent difference between US corporate sites and other international sites was that legal webpages were more frequent in the former. It should therefore be mentioned that the values included in Table 3 represent the frequency of moves and substeps in the subcorpus of legal web texts, and not in the frequency of appearance in all corporate websites as a whole.

Move and step	Substeps	Frequency Original	Frequency Localised
M1. LEGAL DISCLAIMERS			
S1-1. Introduction	a. Welcome	0.00*	17.39
	b. Please read text	13.21	65.22
S1-2. Acceptance legal terms	a. Acceptance	58.49	93.48
	b. Leave website	7.55	15.22
S1-3. Company registration	a. Company legal registration	73.58	34.78
	a. Corporate Address	30.19	30.43
	b. Spanish Tax ID number (CIF)	52.83	30.43
S1-4. Applicable law and jurisdiction		18.87	65.22
S1-5. Copyright- Protected material		67.92	89.13
	a. Written authorisation	26.42	47.83
	b. Which material is protected	26.42	50.00
S1-6. Where is the information stored		0.00*	4.35
S1-7. Website owner...		56.60	71.74
S1-8. Who is the website addressed at?		5.66	23.91
S1-9. Using registered trademarks		3.77	32.61
S1-10. Effective date or revision date		0.00	23.91
M2. PRIVACY POLICY			
S2-1. Compliance to Spanish Privacy Laws			
	a. Spanish Personal Data Protection Law	49.06	13.04
	b. Law 34/2002 of Information Society Services and E-Commerce	22.64	6.52
S2-2. Collection of data		47.17	45.65
S2-3. Right to access, rectify, cancel or oppose data		47.17	28.26
S2-4. Links to and from external websites		43.40	86.96
S2-5. Use of personal data		41.51	36.96
S2-6. Contacting		39.62	58.70

S2-7. Cookies		32.08	50.00
	a. Definition	22.64	19.91
S2-8. Use by Third Parties		26.42	28.26
S2-9. Security		26.42	36.96
	b. Risks	9.43	10.87
S2-10. Notification of use by Third Parties		13.21	4.35
S2-11. Minor Protection		7.55	39.16
S2-11. IP Addresses		5.66	15.22
M3. TERMS AND CONDITIONS			
S3-1. Limitation of liability		67.92	86.92
	a. Problems caused by viruses	41.51	58.70
	b. Errors in site – service interruption	35.85	56.52
	c. Suing for damages	24.53	52.17
	d. Correctness of information	20.75	23.91
	e. Third Party or users' content	5.66	13.04
S3-2. Changes in content		41.51	54.35
S3-3. Changes in terms and conditions		24.53	76.09
S3-4. Appropriate use		22.64	17.39
S3-5. Personal and Private use		16.98	15.22
S3-6. Access restrictions		13.21	17.39
S3-7. Publication of illegal, polemic, pornographic or threatening materials.		9.43	39.13
S3-8. Using site for illicit, illegal, negligent or fraudulent purposes		7.55	2.17
S3-9. Exclusion of warranties		5.66	30.43
S3-10. Printed texts prevail over web texts		5.66	4.35
S3-11. Nielsen ratings		3.77	0.00
S3-13. Sent material is public		1.89	26.09
S3-13. Future expectations		0.00*	10.87

Table 3. Contrastive analysis of moves, steps and substeps in the legal section of corporate websites.

The contrastive study reveals striking differences in the prototypical macrostructures of legal texts in websites. The largest differences are due to a number of steps and substeps that appear with much higher frequencies in LT, together with some substeps that only appear in these last texts. As a whole, the textual block with the largest frequency difference is the substep that encourages users to read the entire legal text before using the site (S1-1-b), +52.1%, followed by a step that refers to changes and modifications in the terms and conditions (S3-4), +51.54%, the links to and from external sites (S2-4), +43.56, and the applicable laws in case of conflict (S1-5), +42%. The higher frequency of the first of the substeps might simply be due to the much larger extension of LT as reported in the previous section. As such, there is an additional need to implicitly encourage the user to read these legal terms. Also of interest is the much higher frequency of the move that established the applicable laws (S1-5). This is clearly indicative of the need to establish implicitly the applicable laws given the high costs associated with the adaptation of legal terms to each target locale. A closer analysis of this step shows that 34% of localised sites do not indicate the applicable legislation, 32.60% chose the Spanish legislation, 21.73% that of the United States, and 8.69% the Swiss one.

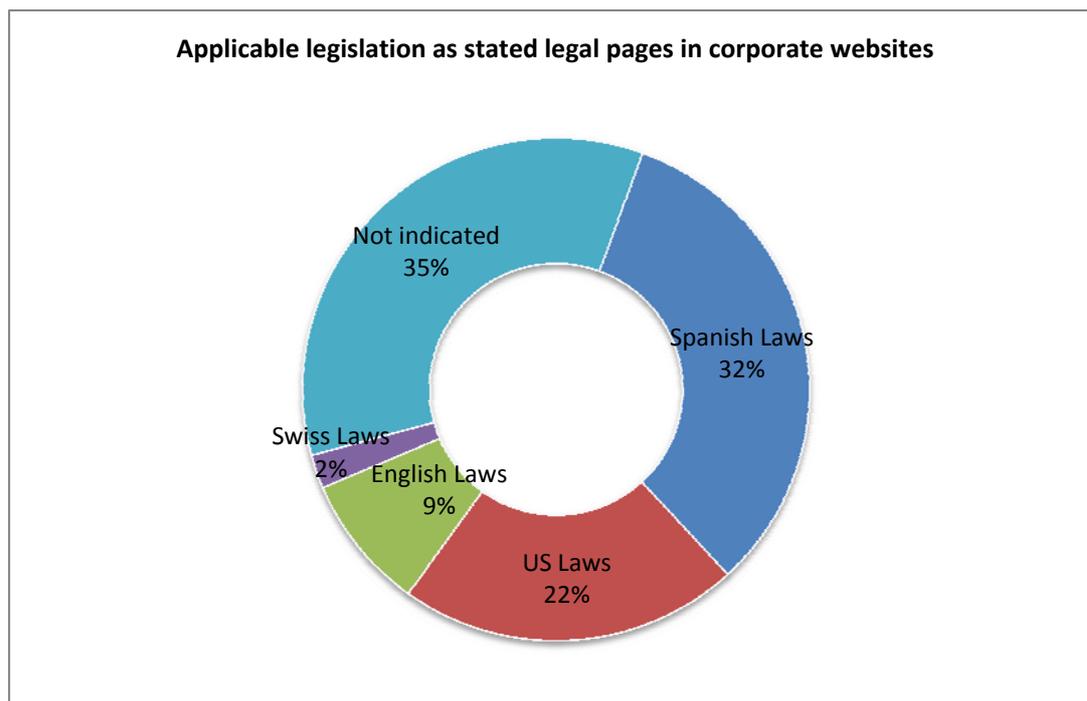


Figure 2. Applicable laws to corporate websites as stated in legal sections

Given that according to the Forbes list all localised companies are based in the USA, it is of interest that in their globalisation-localisation efforts, only 32.60% of websites adapt their legal terms to the target Spanish locale, while 21.73% apply the laws of the USA. Additionally, once the US laws are stated, most sites further delimit the applicable state or federal laws. The following jurisdictions were found, from the most frequent to the least: California, Illinois, Delaware, Washington, Georgia, Minnesota, and US Federal Laws.

As far as the research question, whether websites, or more specifically their legal texts, are adapted to the target locale, the analysis of the segments that are more frequent in original texts is quite revealing. Table 4 shows this contrastive analysis, in which the values are the difference in frequency between both macrostructural profiles presented in Table 3. The five substeps with higher frequencies in original texts are all related to the Spanish legal context, specifically, to the Spanish Laws regarding data protection and e-commerce. The reference to compliance with the Spanish Data Protection Law of 1999 is 38.80% more frequent in original texts, and a specific clause in this law regarding the right of the user to access, rectify or delete this information is also 18.91% more frequent. The substep related to the 2002 Spanish Law of Information Society Services and e-Commerce also appears with a higher frequency, 16.12%. Furthermore, the other two substeps with the larger frequencies in OT refer to specific aspects of business law in Spain: the company legal address and their Spanish Tax Identification Code (CIF), the de facto business ID number in that country.

Original to localised texts	% difference	Localised to original texts	% difference
-Legal registration of company	-38.80%	-Please read the text (S1-1-b)	+52.01%
-Spanish Personal Data Protection Law	-36.01%	-Changes to terms (S1-4)	+51.56%
-Spanish Tax Identification Code (CIF)	-22.40%	-Links from and to the site (S2-4)	+43.56%
-Right to access, rectify, cancel or oppose data	-18.91%	-Applicable Laws (S1-5)	+42.00%
-Spanish Law 34/2002 (Information Society Services and e-Commerce)	-16.12%	-Acceptance of legal terms (S1-2)	+34.99%
		-Minor protection (S2-11)	+31.58%

Table 4. Contrastive analysis of steps and substeps with the greatest intertextual differences in frequency.

It is also of interest to analyse the steps and substeps that do not appear in original legal texts. This might be a probe into source text structures that are not conventional for the same genre in the target culture. Those with the largest differences in frequency are: the substep that welcomes the user to the legal text (S1-1-a), 17%, and the step that indicates the date of revision or effective date of the legal terms (S1-10), 23%. The difference in S1-1a is clearly indicative of a discursive strategy of source English texts that do not exist in the same genre in the target context. In the localised texts, normally the expression *dar la bienvenida* 'welcome' is used, with some potential variation such as *[Company] se alegra de que Ud. Visite esta Página Web* '[Company] is glad that you visit our webpage'. This finding is indicative of a process of localisation in which there is a tendency to maintain the surface structure of the text, regardless of differences in the prototypical macrostructure of the similar genre in the target context. This is consistent with what Larose (1998) refers to as *cloned texts*, that is, translated texts whose macrostructure is fully maintained in the target text regardless of intercultural differences for the same genre. This is also a common effect of localised texts using translation memory software (Jiménez-Crespo 2009b), and it has been shown that maintaining source macrostructures might have an impact in the appreciation of translation quality by end users (Nobs 2006).

Table 4 thus shows that website macrostructures are often fixed and localisers/translators normally do not perform structural changes, even when some communicative blocks might not be relevant for the target audience. It should be noted that any changes to legal provisions in corporate websites are normally carried out during the internationalisation stage that precedes the actual localisation process (Esselink 2000; LISA 2007). In this stage, a review of the literature shows that texts can either be adapted by legal experts, or directly produced in a non culture-specific international form. Consequently, any changes to the provisions during localisation processes need to be informed by legal experts, thus requiring an additional economic investment. The former entails a level of localisation that is referred to as *localized* or *extensively localized* by Signh and Pereira (2005) or *incremental* or *exhaustive localization* by Yunker

(2003: 128-130). The analysis shown in Table 3 demonstrates that, on average, website localisation is far from the ideal level of localisation in which the website is recreated or fully adapted for each target locale, a level that is called *culturally adapted localization* (Sighn and Pereira 2005) or *adapted localization* (Yunker 2003).

5.1.1. A description of the conventional macrostructure of original and localised web legal texts

As previously mentioned, the prototypical macrostructure of any textual genre is normally conventionalised to some extent. If we accept the proposal of 50% of frequency in order to consider conventional any textual feature in any given genre (Gamero 2001; Nielsen 2004),⁸ the steps or substeps with a frequency higher than 50% would represent the conventionalised macrostructure of legal sections in original and localised corporate websites. Table 5 shows these different conventionalised macrostructures. The prototypical macrostructure of OT includes four steps and three substeps, while the LT includes nine steps and six substeps. The steps and substeps that appear more often in localised sites are marked with an asterisk. These probably reflect the most conventionalised steps and substeps in source English texts.

Prototypical macrostructure Original legal web texts	%	Prototypical macrostructure Localised web texts	%
S1. LEGAL DISCLAIMER		S1. LEGAL DISCLAIMER	
S1-2-a. Acceptance of legal terms	58.49	S-1- 1-b. Please read the text*	65.22
S1-3. Legal registration	73.58	S1-2-a Acceptance of legal terms	93.48
S1-3-b. Spanish Tax ID number	52.83	S-1-4. Applicable Law*	65.22
S1-6. Copyright	67.92	S1-6. Copyright	89.13
S1-7. Website owner	56.60	S1-6-a. What material is protected*	50.00
		S1-7. Website owner	71.74
S2. Privacy Policy		S2.Privacy Policy	
S2-1a. Spanish Personal Data Protection Law	49.06	S-2-4. Links to and from external sites *	86.96
		S2-7. Cookies*	50.00
		S-2-6. Contact*	58.70
S3. Terms and conditions		S3. Terms and conditions	
S3-2. Limitation of liability	67.92	S3-2. Limitation of liability	86.92
		S3-2-a. Limitation of liability (Virus)*	58.70
		S3-2-b. Limitation of liability (errors-interruptions in service)*	56.52
		S3-2-c. Limitation of liability (damages)*	52.17
		S3-1. Changes in content*	54.35
		S3-3. Changes in terms and conditions*	76.09

Table 5. Prototypical macrostructure of Spanish original and localised web legal sections in corporate websites.

As can be observed in the contrastive analysis, these macrostructures are clearly distinct. Most of the differences are concentrated in the moves *privacy policy* and *terms and conditions*. The macrostructural differences can be classified in three distinct categories: those due to (1) discursive strategies in the source texts that do not appear in the similar genre in

the target context, such as the step *welcome* or *please read the text*, (2) steps that appear as a consequence of differences in the source and target legal contexts, such as minor protection, changes and acceptance of legal terms in LT or the steps related to the Spanish business IDs, Spanish privacy and e-commerce laws, etc., (3) the localised nature of websites that shapes a step in which the legislation that would apply in case of conflict is implicitly established. These differences indicate that, despite the industry efforts to make localised texts *look* like those originally produced in-country (LISA 2004), legal texts clearly show different prototypical macrostructures. As a result, this raises the question of whether localised websites could be considered a specific or parallel digital genre of its own (Shreve 2006; Jiménez-Crespo 2010).

5.2. Phraseological analysis associated to rhetorical moves

Given that the macrostructural differences have been explored, this section focuses on the impact of macrostructural difference in the microstructures or actual language used. In order to contrastively analyze the phraseology chosen by localisers and given the limitations of this paper, four representative steps and their associated conventional linguistic units were selected. These are *acceptance of legal terms* (S1-2-a), *website owned by...* (S1-7), *limitation of liability* (S3-2), and *limitation of liability due to errors* (S3-2-b). For each of them a node word associated to the step was selected and a collocation analysis was performed in each section of the corpus.

5.2.1. Acceptance of legal terms (S1-2-a)

The first contrastive phraseological analysis was performed in the step in which the terms of the website are accepted. The node chosen for the concordance analysis was the lemma *accept**. It was observed that OT texts favored the use of constructions with the noun form *aceptación* 'acceptance', and therefore, the analysis concentrated on the collocations of this noun. The analysis clearly shows that this noun appears mostly as the main node in two clusters *aceptación plena y sin reservas* or *aceptación sin reservas*. Only one potential variation was found, in which the adjective *plena* was substituted for *expresa*. This cluster is mostly preceded by the verb *implicar*, and to a lesser degree, *suponer*.

N	Concordance
1	usuario expresa su comprensión y aceptación plena y sin reservas
2	delante, el "Usuario") e implica la aceptación plena y sin reservas
3	El mero uso del Portal implica la aceptación sin reservas por parte
4	. El mero uso del Portal implica la aceptación sin reservas por parte
5	El acceso a dicho sitio implica su aceptación sin reservas. La utilizac
6	El acceso al mismo implica su aceptación sin reservas. • 3. S
7	et. El acceso al mismo implica su aceptación sin reservas. La utiliza
8	en el área de clientes implicará su aceptación expresa y sin reserva
9	n de la Tarjeta A.G.C. supondrá la aceptación plena y sin reservas,
10	n de la Tarjeta A.G.C. supondrá la aceptación plena y sin reservas,
11	. La visita del sitio web supone la aceptación plena y sin reservas

Figure 3. Concordance analysis of the noun *aceptación* in the original legal corpus that illustrates its conventionalised phraseological forms.

A similar analysis using the same noun in LT yielded a lower frequency of use with only six occurrences. The most conventional form in original texts, *la aceptación plena y sin reservas* appears in two cases (33.3%), and the same form without the adjective *plena* also appears twice (33.3%). Two cases show variations of the same for that are inexistent in the original corpus, such as *aceptación, sin reserva alguna*, and *aceptación sin limitaciones*. It is also indicative that two of the concordance lines use commas for adjectives that follow *aceptación*, and a verb that is inexistent in original texts appears in localised texts, *manifestar*, in concordance lines 3 and 4.

N	Concordance	Set Tag
1	inados servicios implicar. Asimismo la aceptación , sin reserva alguna, de las	
2	(en adelante, el "Usuario") e implica la aceptación plena y sin reservas de tod	
3	b, el visitante ("Usuario") manifiesta su aceptación sin reservas de las condi	
4	la página web del Usuario manifiesta su aceptación sin reservas de las present	
5	vegación por este sitio Web supone su aceptación sin limitaciones de estos t	
6	xpress.es" implica el conocimiento y la aceptación , plena y sin reservas, de lo	

Figure 4. Concordance analysis of the noun *aceptación* in the localised subcorpus.

5.2.2. Website owned by... (S1-7)

The following substep under study informs the user who is the legal owner of the website. In the case of Spain, all websites need to inform the user of the legal person responsible for any website under the Law 34/2002 of Information Society Services and E-Commerce. In this case, a manual analysis was performed on all the phraseological units used in this substep were recorded. The most frequent construction in original Spanish sites is *[Company] es titular de [website]* '[company] owns the website' with a frequency of 40%, followed by *[Company] pone a disposición de los usuarios de Internet*, '[company] provides all Internet users with [the website]', in 25% of OT. The former phraseological unit has a frequency of 4.34% in LT, and the latter does not appear at all. These two most frequent collocations in OT replicate the language used in current Spanish legislation, a common intertextuality effect in legal texts.

% Original	Phraseological unit	% Localised
40%	...es titular de...	4.34%
25%	[empresa] pone a disposición de los usuarios de Internet ...	0%
15%	...en cumplimiento de la Ley ...	4.34%
15%	...es propiedad de...	26.08%
5%	...fue creado por ...	4.34%
0%	...es operado ...	17.39%
0%	...gestiona-es gestionado...	13.04%
0%	...mantiene...	13.04%
0%	...es de...	8.69%
0%	...es administrado...	4.34%
0%	[Empresa] dirige el sitio...	4.34%

Table 6. Contrastive phraseological analysis of forms associated to the substep S-1-7, Website owner.

In LT the most frequent collocation is *[site] es propiedad de [company]* ‘[site] is owned by [company]’, with 26.08% of instances. While in original texts only five different constructions are used, LT shows a much wider range of variation, with eleven constructions found. The most interesting aspect is the fact that the combined frequency of constructions that do not appear in original Spanish texts is 60.84%. This finding clearly shows that even when this structural substep is very frequent in both original (56.60%) and localised texts (71.74%), the latter do not show similar frequencies or patterns of phraseological use. In fact, it can be observed that many of the patterns that appear only in localised texts could be considered lexical and syntactic anglicisms, such as *[the site] es operado por ...* ‘[the site] is operated by ...’, *... mantiene [the site]* ‘... maintains [the website]’ or *[the site] es administrado por X* ‘[the site] is administered by...’.

5.2.3. Limitation of liability (S3-2)

This step and its constituent substeps are among the most frequent textual blocks in all legal texts. The goal of these sections is to limit the liability and responsibility of the company in case problems or errors arise while, or as a result of, using the website. The most frequent lemma in original texts was *responsabl**, given that these texts list all the circumstances under which the company should not be held responsible. The most frequent constructions in both corpora show similar patterns of use. In OT, the most frequent constructions that correspond to variations of the English phrases ‘will not be liable – will not be responsible’ are: *...no se hace responsable* (OT=25.8%, LT=23.5%), *... no será responsable...* (OT=23.5.8%, LT=21.6%), and *...no se responsabiliza...* (OT=21.1%, LT=26.4%). In both cases, the lemma *responsabl** is mostly used in its adjectival form (OT=59.8%, LT=60.92%), while there is a slight difference in the use of its verbal (OT=21.1%, LT=9.98%) and noun forms (OT=19.1%, LT=29.1%).

Collocation	Freq. Original OT	Freq. Localized, LT
...no se hace responsable de ...	25.80%	23.50%
...no será responsable de ...	23.50%	21.60%
...no se responsabiliza de ...	21.10%	26.40%
...no asume responsabilidad alguna por ...	11.70%	16.98%
...no es responsable de cualquier...	10.50%	11.32%
...no tiene responsabilidad ...	1.10%	4.70%
...se libera de toda responsabilidad ...	1.10%	0%
...no adquiere responsabilidad ...	1.10%	0%
...no acepta responsabilidad alguna...	1.10%	0%
...rechaza la responsabilidad sobre...	1.10%	0%
...queda exonerada de responsabilidad ...	1.10%	0%
Additional clusters that only appear in localised texts (Frequency in LT=0.9%)		
no está sujeto a responsabilidad no admitirá responsabilidad legal no asumirá responsabilidad no asumimos responsabilidad no puede asumir responsabilidad no podrá ser considerado responsable no nos hacemos responsables de acuerdo no hacer responsable límite de la responsabilidad límite a la responsabilidad exclusión de la responsabilidad limitación de la responsabilidad	no asume ninguna responsabilidad no tendrá responsabilidad <u>mantiene la no responsabilidad</u> no nos responsabilizamos no podrá responsabilizarse a no se responsabilizará acepta que no sean responsables no somos responsables será su responsabilidad no tendrá responsabilidad no haya incurrido en una responsabilidad	

Table 7. Contrastive analysis of the lemma *responsab in OT and LT.**

The greatest variation in both sections of the corpus is in the verbs that collocate with the noun *responsabilidad* ‘responsability’. In any case, the most interesting aspect is that LT show a much greater range of variation in its potential collocations, with 23 different forms (20.8%) that do not appear in OT. This greater variation found in this and the previous analysis is consistent with findings in other sections of localised corporate websites when compared to original Spanish websites (Jiménez-Crespo 2009a).

5.2.4. Limitation of liability due to errors (S3-2-b).

In this step and its substeps, the company responsible for the website intends to limit any responsibility due to potential errors in the website. The most frequent collocations of the node *error* are *typographic*, *spelling*, *content*, *omissions*, *security* and *translation*. The last collocation is of great interest to our study given that the company implicitly indicates that the website is localised. This type of error is mentioned in 6.52% of the sites included in the subcorpus. Normally, localised versions are presented in legal texts as an additional service that the company provides for the convenience of the user and, as a consequence, in case of legal disputes, only the source English version would prevail. The following segment from the UPS websites illustrates this issue:

Las traducciones a otros idiomas, en caso de haberlas, se facilitarán únicamente a efectos prácticos, y la versión inglesa podrá visionarse de las siguientes maneras. 'Translations into other language, in case they exist, are provided exclusively for practical purposes, and the source English version can be viewed the following way.'

The most interesting phraseological construction in this substep is the translation of the English collocation *"as is"*, always inserted in quotation in LT. This collocation is normally used to indicate that the contents or the website is offered *"as is"*, with no written and express warranty. This is a recurrent phraseological unit in source English texts that, nevertheless, does not appear in Spanish OT. However, the direct translation of this conventionalised unit appears in 34.78% of LT, almost always in the quotation marks. Figure 5 shows a concordance analysis in the localised corpus using as a node the word *tal* 'as'.

N	Concordance
9	lementos (sin carácter limitativo) se proporcionan "TAL CUAL" y están 'DISPONIBLES' sin manifes
10	cionamiento, y acepta que el sitio se ofrece "tal cual". Sealed Air Corporation no ofrece ningú
11	RAMENTE INFORMATIVOS Y SE PRESENTAN "TAL CUAL". POR CONSIGUIENTE, AL OFREC
12	TENIDOS DE ESTE SITIO WEB SE OFRECEN "TAL CUAL". MEDIANTE ESTE DOCUMENTO,
13	Garantías* EL SERVICIO SE PROPORCIONA "TAL CUAL". ADM NO OFRECE NINGÚN TIPO
14	Í CONTENIDOS SE PROPORCIONAN EN BASE "TAL CUAL", SIN GARANTÍAS DE NINGUNA CL
15	DUCTOS DE TERCEROS, SE PROPORCIONAN "TAL CUAL" Y "SEGÚN DISPONIBILIDAD" PAR
16	su contenido. Los juegos en fase beta se ofrecen "tal cual" y "según disponibilidad" y no le proporc
17	te sitio es responsabilidad suya. El sitio se ofrece "tal cual" y "según disponibilidad". Nos reservam
18	abilidad LOS MATERIALES SE SUMINISTRAN "TAL CUAL" SIN NINGÚN TIPO DE GARANTÍA,
19	FRECEN EN ESTE SITIO SE PROPORCIONAN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO,
20	ADOS CON LOS MISMOS SE PROPORCIONAN "TAL CUAL" SIN GARANTÍA DE NINGÚN TIPO.
21	sable de los mismos, los cuales se proporcionan "tal cual" ('as is') sin garantías de ningún tipo, ex
22	ro no está limitada a, la información de contacto (tal como su nombre, dirección postal, dirección
23	es tipográficos. Todo en este sitio se proporciona "tal como es", sin garantía de ningún tipo, ya se
24	l contenido de este sitio web se ofrece al usuario "TAL COMO ESTÁ", SIN GARANTÍA DE NIN

Figure 5. Concordance analysis of translations of the collocation *"as is"* using the word *tal* 'as' as a node in localised texts.

It can be clearly observed that the source text conventions have been transferred to the target texts. In concordance line 21, the English form is even kept next to the translation, a translation strategy that signals the difficulty when dealing with highly conventionalised forms in source texts that do not have a counterpart in target texts. The use of quotation marks in legal texts does not appear in original Spanish texts, and therefore, it could be labeled as a typographic anglicism (Martinez de Sousa 2007). Some other possible translations of these collocations were also found through the use of quotation marks, such as:

- **"COMO SE ENCUENTRA", "COMO ESTÉ DISPONIBLE[sic]"** [Purina]
- **"en el estado en que se encuentran"** [Genworth Financal]
- SON PROPORCIONADAS EN SU **"ESTADO ACTUAL[sic]"** [Carterpillar]
- Sin limitar lo precedente, todo en el Sitio Web le es proporcionado en **el estado en que se encuentra** [mattel.es]
- **"EN LAS CONDICIONES QUE SE ENCUENTRAN[sic]"** [HP.es]

Again, it can be observed that localised texts resort to a wide range of phraseological variation. Even when, as seen in Figure 5, the most frequent Spanish translation is *"tal cual"*, seven other possible translations were also identified by tracing quotation marks in the corpus. This semantic and typographic anglicism clearly indicates that the microstructures of original texts are replicated in the translations, even when the resulting rendering might not find any counterpart in similar texts directly produced in the target language. From a practical perspective, the genre-based approach taken can be highly beneficial to practitioners that have to deal with this type of texts, as descriptive macrostructural studies can help identify how original texts formulate the same communicative purpose.

6. Conclusions

The goals of the localisation industry for the localisation process are to produce quality websites that "look like they have been developed in country" (LISA 2004: 11), as well as to offer users the most credible and usable localisations. With this in mind, the purposes of this study were twofold. On the one hand, it intended to research the *adaptation* by the localisation industry in order to observe to what extent localised corporate websites are adapted to the target sociocultural contexts. On the other hand, this paper intended to research contrastively the macrostructural and phraseological differences between original and localised legal web texts, a direct probe into the translation strategies adopted both by the commissioners or initiators and by the translators/localisers themselves. As far as the first goal is concerned, it has been observed that only 32.60% of websites with legal content adapt to some degree their legal terms to the Spanish legal system, while despite the fact that all companies responsible for the localised websites are based in the USA, only 21.73% implicitly mentioned that the laws and courts of the United States should apply. Nevertheless, it has been observed that the *adaptation* of legal texts cannot be simply established through the adaptation to local legislation, as additionally, macrostructures and microstructures also play an essential role if websites are to be received as *local productions*. In fact, the study identified remarkable differences in the macrostructures and microstructures used if LT are compared to similar texts directly produced in Spanish. The macrostructural differences observed were classified either as a result of (1) different discursive strategies in the source and target context, (2) differences in the legal systems involved, or (3) the internationalised non-culture specific legal text having been developed for an international audience. In the latter case, despite the fact that the company localised the site to better serve the target audience, the localisation strategy adopted entails an implicit mention of the lack of legal validity of the localised version. Despite the fact that legal texts are rarely fully read by users (Price and Price 2002),

this raises interesting questions about the view in the industry about the role of localisations and the socioprofessional status of localisers.

It has also been observed that the differences in the same textual genre in both source and target contexts have an impact on phraseology and terminology. In some cases, translators might not have any ready-made conventionalised form to use in the target language. Further, despite the existence of conventionalised forms, translators might simply replicate source language structures such as shown by the phrase *"as is"*. It is of interest that, on average, LT shows greater phraseological variation than original texts, an effect already observed in localised web texts (Jiménez Crespo 2009a, Jiménez-Crespo 2009b). This means that in highly conventionalised web genres, the localisation process might lead to texts with higher levels of phraseological and terminological variation than similar texts directly produced in the target language. As an example, the case of the step *website owned by* showed that 60% of renderings in localised texts were not found in original Spanish texts, among which several were inappropriate syntactic anglicisms.

As far as the implications for the practice and training in web localisation, it should be mentioned that despite advanced technological competence, web localisation is an extremely complex process in which many translation types and modalities can be found. Though mostly a case of instrumental translation, legal web texts entail documentary translations and this proves that localisers require advanced translation competencies in a wide range of translation types. In this regard, this study has shed some light onto the many components of the scarcely researched notion of *localization competence* (Jiménez and Tercedor, forthcoming). Normally, the predominant industry perspective on localisation training assumes that the core of localisation training entails proficient use of all types of technology tools and understanding of technological processes and programming languages. Nevertheless, as witnessed by the results of this study, the localisation process of any website entails a complex mix of subtexts related to several translation types, such as marketing, technical, multimedia, economic, instructional legal, journalistic, etc. (Jiménez 2008b; 2009b). As such, the level of translation competence required to deal with such a varied group of texts is very advanced.

Last but not least, the localisation process is inevitably immersed in a global cycle with strict money and time constraints (Wright 2006), and therefore, the objectives or goals of the commissioners or initiators have to be conceptualised as a relatively unattainable goal. In fact, the differences between legal OT and LT can be attributed to both constraints on the part of the commissioners that require a specific localisation level, and the constraints of the localisation process itself that, the same as translation, is "a communicative event which is shaped by its own goals, pressures and context of production" (Baker 1996:175). Web localization

is still an underresearched area in translation studies, and it is hoped that this paper will contribute to both the practice of web localisation and the theoretical conceptualisations of this fascinating and ever-increasing phenomenon.

Bibliography

- **American Bar Association** (2007). *The In-house counsel's essential toolkit / ABA Committee on Corporate Counsel*. Chicago: Section of Business Law, American Bar Association.
- **Austermühl, Frank** (2006). "Training Translators to Localize." Anthony Pym, Aleksander Perekstenko and Bram Starink (Eds) (2006). *Translation Technology and its Teaching*. Tarragona: Intercultural Studies Group, 69-81.
- **Baker, Mona** (1995). "Corpora in Translation Studies: An overview and some suggestions for future research." *Target* 7, 223-243.
- — (1996). "Corpus-based Translation Studies: The Challenges that Lie Ahead." Harold Somers (ed) (1996). *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamins, 175-186.
- — (1993). "Corpus Linguistics and Translation Studies: Implications and Applications." Mona Baker, Francis Gill and Elena Tognini-Bonelli (Eds) (1993). *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233-250.
- **Bhatia, Vijay K.** (1993). *Analysing genre. Language use in professional settings*. London: Longman.
- **Bly, Robert** (2002). *The Online Copywriter's Handbook*. New York: McGraw-Hill.
- **Borja Albí, Anabel** (2005). "Organización del conocimiento para la traducción jurídica a través de sistemas expertos basados en el concepto de género textual." Isabel García Izquierdo (ed.) (2005). *El género textual y la traducción. Reflexiones teóricas y aplicaciones pedagógicas*. Berna: Peter Lang, 37-67.
- — (2000). *El texto jurídico inglés y su traducción al español*. Barcelona: Ariel.
- — (2007). "Corpora for Translators in Spain. The CDJ-GITRAD Corpus and the GENTT Project." Gunilla Anderman and Margaret Rogers (Eds) (2007). *Incorporating Corpora. The Linguist and the Translator*. Clevedon: Multilingual Matters, 243-65.
- **Cao, Deborah** (2007). *Translation Law*. Clevedon/Buffalo/Toronto: Multilingual Matters
- **Codina, Lluís** (2003). "Hiperdocumentos: composición, estructura y evaluación." Javier Díaz Noci y Ramon Salaverría Aliaga (Eds) (2003). *Manual de redacción ciberperiodística*. Barcelona: Ariel Comunicación, 141-194.

- **Comité sectoriel** de l'industrie canadienne de la traduction (1999). *L'industrie canadienne de la traduction. Stratégie de développement des ressources humaines et d'exportation*. Ottawa: Comité sectoriel de l'industrie canadienne de la traduction
- **Cronin, Michael** (2003). *Translation and Globalization*. London: Routledge
- **Crystal, David** (2001). *Language and the Internet*. Cambridge: Cambridge University Press.
- **Dunne, Keiran** (2006). "A Copernican Revolution." Keiran Dune (ed.) (2006). *Perspectives on localization*. Amsterdam and Philadelphia: John Benjamins, 1-11.
- **Esselink, Bert** (2000). *A Practical Guide to Localization*. Amsterdam/Philadelphia: John Benjamins.
- **Fritz, Gerd** (1998). "Coherence in hypertext." Wolfram Bublitz, Uta Lentz and Eija Ventola (Eds) (1998). *Coherence in Spoken and Written Discourse: How to Create it and How to Describe it*. Amsterdam/Philadelphia: John Benjamins, 221-234.
- **Fernández Sánchez, Franciso** (2004). *El folleto de cursos de idiomas para extranjeros: Análisis contrastivo (alemán-español) por tipos de emisor y subtextos*. PhD dissertation. Universitat Pompeu Fabra.
- **Gamero Pérez, Silvia** (2001). *La traducción de textos técnicos*. Barcelona: Ariel.
- **Garrand, Tim** (2001). *Writing for Multimedia and the Web*. Woburn, MA: Butterworth-Heinmann.
- **Gläser, Rosemarie** (1979). *Fachstile des Englischen*. Leipzig: VEB Verlag Enzyklopädie.
- **González Aguilar, Audilio, Carlos Ferreyros Soto and Valentin Carrascosa López** (2004). *Los contratos en la sociedad de la información. Formularios de contratos informáticos e Internet*. Granada: Comares.
- **Göpferich, Susanne** (1995). *Textsorten in Naturwissenschaften und Technik. Pragmatische Typologie- Kontrastierung-Translation*. Tübingen: Gunter Narr.
- **Gouadec, Daniel** (2007). *Translation as a profession*. Amsterdam/Philadelphia: John Benjamins.
- **Kelly, Kendall** and Jennifer Jones. 2007. "Websites and the Law: an Avenue for Localization." *Linguistic and Culture Issues in the Online Global Environment*. Kirk St. Amant, ed. Hershey, PA: Idea Publishing Group, 202-14
- **Hanks, William F.** (1996). *Language and Communicative Practices. Critical Essays in Anthropology*. Oxford: Westview Press.
- **Holmes, James** (1988). "The Name and Nature of Translation Studies." James Holmes (ed.) (1988). *Translated! Papers on Literary Translation and Translation Studies*. Amsterdam: Rodopi, 67-80.
- **Hurtado Albir, Amparo** (2001). *Traducción y Traductología. Introducción a la traductología*. Madrid: Cátedra.
- **Jeney, Cynthia** (2007). *Writing for the Web: a Practical Guide*. Columbus, OH: Pearson Prentice Hall.

- **Jiménez-Crespo, Miguel A.** (2008a). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. PhD Dissertation. University of Granada.
- (2008b). "Caracterización del género 'sitio web corporativo' español: análisis descriptivo con fines traductológicos." Maria M. Fernández Sánchez and Ricardo Muñoz Martín (Eds) (2008). *Aproximaciones cognitivas al estudio de la traducción e interpretación*. Granada: Comares, 259-300.
- — (2009a). "Conventions in localization: a corpus study of original vs. translated web texts." *JoSTrans. The Journal of Specialised Translation* 12, 79-102.
- — (2009b). "The effect of Translation Memory tools in translated web texts: evidence from a comparative product-based study." *Linguistica Antverpiensia*, 8, 213-232.
- — (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.
- Jiménez-Crespo, Miguel A. and Tercedor, Maribel (forthcoming). "Applying corpus data to define needs in web localization training". *Meta*.
- Larose, Robert (1998). "Méthodologie de l'évaluation des traductions." *Meta* 43, 163-186.
- Laviosa, Sara (1998). "The Corpus-based Approach: A New Paradigm in Translation Studies." *Meta* 43 (4), 474-479.
- LISA. (2004). *Localization Industry Primer, 2nd Edition*. Geneva: The Localization Industry Standards Association (LISA).
- —(2007). *LISA Globalization Industry Primer*. Romainmôtier, Switzerland: LISA.
- Liu, Chang and Kirk. P. Arnett (2000). "An examination of Privacy Policies in Fortune 500 Web Sites." *Mid-American Journal of Business* 17 (1), 13-21.
- Mangiron, Carmen and O'Hagan, Minako (2006). "Game Localisation: Unleashing Imagination with 'Restricted' Translation." *Jostrans* 6, 10-21.
- Martínez de Sousa, Jose (2007). *Manual de estilo de la lengua española MELE 3 (3rd edition)*. Somonte-Cenero: Ediciones Trea.
- Mata Pastor, Manuel (2005). "Localización y traducción de contenido web." Detlef Reineke (ed) (2005). *Traducción y Localización*. La Palmas de Gran Canaria: Anroart Ediciones, 187-252.
- Microsoft Corporation (2003). *Developing International Software*. Redmond: Microsoft Corporation.
- Mazur, Iwona (2007). "The metalanguage of localization: Theory and practice." *Target* 19 (2), 337-357.
- Nielsen, Jacob (1999). *Top ten mistakes revisited*. Alertbox. On line at <http://www.useit.com/alertbox/990502.html> (consulted 00.00.2010)

- **Nielsen, Jacob** and Loranger, Hoa (2006). *Prioritizing Web Usability*. Indianapolis: News Riders.
- **Nielsen Jacob** and Marie Tahir (2002). *Homepage Usability: 50 Websited Deconstructed*. Indianapolis: News Riders.
- **Nobs, Marie L.** (2006). *La traducción de folletos turísticos: ¿Qué calidad demandan los turistas?* Granada: Comares.
- **Nord, Christiane** (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained*. Manchester: St. Jerome.
- **ÖNORM D 1200** (2000). *Translation and Interpretation Services, Translation Services. Requirements of the Service and the Provision of the Service*. Vienna: Österreichisches Normungsinstitut.
- **Paltridge, Brian** (1997). *Genre, Frames and Writing in Research Settings*. Amsterdam and Philadelphia: John Benjamins.
- **Price, John** and Lisa Price (2002). *Hot Text. Web Writing that Works*. Berkeley, CA: News Riders.
- **Pym, Anthony** (2003). "What localization models can learn from Translation Theory." *The LISA Newsletter: Globalization Insider* 12.
- — (2005). "Localization. On its nature, virtues and dangers". *Synaps* 17, 17-25.
- — (2009). *Exploring Translation Theories*. London: Routledge.
- **Reiss, Katharina** and Hans J. Vermeer (1984). *Grundlegung einer Allgemeinen Translationstheorie*. Tübingen: Niemeyer.
- **Robbins, Stephanie** and Antonis Stylianou (2003). "Global corporate web sites: an empirical investigation of content and design." *Information & Management* 40, 205-212.
- **Santini, Marina** (2007). *Automatic Identification of Genre in Webpages*. Doctoral dissertaion. Universidad de Brighton.
- **Schäler, Reinhard** (2005). "The Irish Model in Localization." Paper presented at *LISA Forum Cairo 2005: Perspectives from the Middle East and Africa*. (Cairo, 5-8 December 2005). Online at <http://www.lisa.org/utills/getfile.html?id=61136686> (consulted 5-6- 2009).
- **Shepherd, Michael** and Carolyn Watters (1998) "The evolution of cybergenres." *Proceeding from the XXXI Hawaii International Conference on System Sciences*, edited by Ralph Sprague. Los Alamitos, CA: IEEE-Computer Society, 97-109.
- **Shepherd, Michael**, Carolyn Watters and Alistair Kennedy (2004). "Cybergenre: Automatic Identification of Home Pages on the Web." *Journal of Web Engineering* 3 (3-4), 236-251.
- **Shreve, Gregory M.** (2006). "Corpus Enhancement and localization". Keiran Dunne (ed) (2006). *Perspectives on Localization*. Amsterdam-Philadelphia: John Benjamins, 309-331.

- **Singh, Nitish and Arum Pereira** (2005). *The Culturally Customized Web Site: customizing web sites for the global marketplace*. Oxford: Elsevier.
- **Storrer, Angelika** (2002). "Coherence in text and Hypertext." *Document Design* 3 (2), 157-168.
- **Swales, John M.** (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- **Toury, Gideon** (1995). *Descriptive Translation Studies and Beyond*. Amsterdam and Philadelphia: John Benjamins.
- **Yunker, John** (2003). *Beyond Borders: Web Globalization Strategies*. Indianapolis: New Riders.

Biography

Miguel A. Jiménez-Crespo, PhD, coordinates the Master and BA program in Spanish Translation and Interpreting at Rutgers University, USA. He completed a PhD and MA-BA degree in Translation and Interpreting at the School of Translation and Interpreting, University of Granada, Spain. He has also studied at the University of Glasgow and Moscow State Linguistic University, where he taught Spanish and Translation. His research and publications concentrate on localisation, the translation of digital texts such as software or web content, as well as translation training, terminology and corpus linguistics.



Contact: Miguel Jimenez [miguelfi@rci.rutgers.edu]

¹ In fact, the five most used websites in the world, Google, Microsoft, Yahoo, Facebook and Ebay (Nielsen Netratings 2010), offer a growing number of localised versions of their sites.

² It should be noted that *printed hypertexts* such as encyclopedias or phone books also exist (Fritz 1998).

³ It would be incorrect to identify a *node* exclusively with a web page. A web page is simply the unit of storage and retrieval on the WWW (Nielsen and Loranger 2006), and it can contain several *nodes*.

⁴ This would not be the case of RSS news feeds, as they appeared exclusively on the Internet and are a distinct web digital genre.

⁵ Textual macrostructure can be defined as a conventionalised sequence in which certain textual elements that are thematically and functionally invariable and that occur in a somewhat flexible hierarchical order (Göpferich 1995: 127; Hurtado Albir 2001: 495)

⁶ i. e. OECD privacy policy generator

<http://www.oecd.org/document/42/0,3343,en_2649_34255_28863271_1_1_1_1,00.html>, Freeprivacypolicy <<http://www.freeprivacypolicy.com/>> ,

<<http://javascript.about.com/library/blprivacy.htm>>, Dma privacy policy generator <http://www.the-dma.org/privacy/privacypolicygenerator.shtml>

⁷ Following Gläser (1979: 90), the notion of conventional linguistic form is here defined as a frequent phraseological or lexical unit whose function is to formulate recurring units of meaning that, depending on the textual genre in question, is readily available to the text producer.

⁸ The actual frequency in order to consider any feature conventional in a given genre varies in different publications, such as 70% (Hoffman 1988), or 90% (Fernández Sánchez 2004). From a web usability perspective, Nielsen differentiates between standard features with a frequency of 80% or more of websites, and conventions from 50% to 79%.