JoSTrans
The Journal of Specialised Translation

Conventions in localisation: a corpus study of original vs. translated web texts.

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

ABSTRACT

The crucial role of conventions in translation has been extensively recognised in Translation Studies. The localisation of digital texts entails mainly instrumental texts that need to be received as original productions in the target context of reception, and therefore, source-text conventions should in principle be replaced by those in the same target genre (Nord 1997; Gamero 2001). After a theoretical review of the notion of convention, this paper will consider whether target text conventions are actually incorporated in localised texts. For this purpose, web navigation menus were selected, as they represent the most conventional textual segments in websites (Nielsen & Tahir 2002). For the purpose of this study, a comparable corpus of navigation menus was extracted from the Spanish Comparable Web Corpus (Jiménez-Crespo 2008a). Following a descriptive study on conventional features of original Spanish corporate websites (Jiménez-Crespo 2008b), these findings were contrasted with translated texts in order to shed some light onto the role of convention in the translation products. The results of the contrastive study show that conventional terminology in Spanish websites is significantly less present in translated texts. This is explained in terms of interference from source texts conventions and the specific constraints that operate during the translation process.

1. Introduction: translation, localisation and the digital revolution

In our globalised world, the role of digital texts is more crucial than ever. The rapid evolution of the Internet, the World Wide Web and software products has caused a technological revolution in the way texts are produced, translated and distributed around the world. This revolution would not have been possible without the emergence of a new translation modality, localisation. Localisation is a technological, textual, communicative and cognitive process by which interactive digital texts are adapted for their use in a different cultural and linguistic context of reception. Localisation is a relatively young phenomenon whose roots can be traced back to first attempts at translating software in the 1980s (Esselink 2006). In its early stages, localisation gradually allowed digital texts to reach an increasing international body of users, and as a consequence, this process contributed immensely to the exponential increase of digital content on the web.

The Localisation Industry, a new consolidated economic sector, emerged in the late 1990s in the context of this digital revolution. Nevertheless, the fast pace at which the technology advanced meant that it had to establish itself without fully relying on the body of knowledge of Translation Studies (Dunne 2006a; Folaron 2006). In this sense, it has been argued that the Localisation Industry has not fully benefited from a number of theoretical concepts that could help shape many localisation processes in a more effective way. Among others, this is the case of the notion of convention, the main focus of this paper.

Digital web texts have been used for over two decades around the world and, therefore, the evolution and establishment of conventional features in digital texts can be researched in each locale or linguistic community. At the beginning, software and web texts in different languages were heavily influenced by conventional features in English texts (Schäler 2002). In fact, digital texts have been identified as one of the main vehicles of transmission of conventions from the dominant English language around the world (House 2001). However, the increasing production and use of digital texts by different linguistic discourse communities may have led to the emergence of distinctive conventions. In fact, one of the basic principles in convention is that the process by which they are established is random and consequently, most linguistic communities might not share the same conventions for similar textual genres (Göpferich 1995; Nord 1997). This principle also applies to digital texts. Therefore, it would be inadequate to assume that conventional features in English texts that were directly transferred to most target languages (such as the calque of the lexical units 'contact us' or 'about us' in web navigation menus) are at this point conventional in most locales.

Localisation has also benefited from the advances in the fields of usability and web development. Researchers in these disciplines have focused on the role of convention in human-computer interaction, and in fact, one of the primary recommendations is to follow established conventions (Vaughan & Dillon 2006; Nielsen & Loranger 2007; Nielsen 2004; Brink et al. 2002). Compliance with established conventions during web development has been directly linked to its future success (Nielsen 2004). Given the importance of this notion in digital texts, the objective of this paper is to explore the role of convention in the localisation process. More specifically, this paper aims to study the extent to which the localisation process produces texts that incorporate conventional features established in the target locale. In so doing, the question of whether web texts are influenced by genre conventions from source texts will be explored (Reiss & Vermeer 1984). Thus, after reviewing the concept of 'convention' from the perspective of the Localisation Industry and Translation Studies, an empirical study of terminological conventions will be carried out. The analysis will be based on a comparable corpus of original and localised navigation menus from corporate websites. The methodology is based on

identifying and recording the frequency of all terms that denote the same structural block in websites, such as 'contact us,' 'history' or 'privacy policy.' Conventions can be objectively quantified through frequency of use, and therefore the analysis will use this variable in order to contrast the conventionality level of all terms related to each structural concept. This will allow to observe whether both textual populations share the same terminological conventions.

2. Theoretical framework: conventions and the objective of localisation

During recent years the localisation process has been the object of an increasing number of studies (Jiménez-Crespo 2008a: 68-69). From a translation perspective, it should be mentioned that many scholars have argued that the Localisation Industry and Translation Studies should combine their efforts towards a better understanding of the this phenomenon (Gouadec 2003; Pym 2004; Dunne 2006a; Folaron 2006; Mazur 2008; Jiménez-Crespo 2008a). This paper asserts that this goal could be achieved if this new translation modality is conceptualised in terms of functional approaches to translation (Reiss and Vermeer 1984; Holz-Mänttäri 1984; Nord 1997). Functional approaches to translation assign a special role to the *skopos* of the translation, that is, its function in the context of reception. This function of the translation can usually be summarised in the "translation brief" (Nord 1997), a compendium of the function of the translation and its context of reception. Normally, in those cases in which initiators or commissioners might fail to provide this information, professional translators are expected to deduce it based on their expert knowledge. This information is essential in order to produce a fully functional target text. However, even when initiators might provide this information, this paper agrees with Dunne (2006b: 100) in that "[c]lients often *cannot* provide all the necessary standards, requirements or specifications for the simple reason that they are unfamiliar with the languages, culture, conventions and legal requirements of the target locale(s)". Therefore, how can localised texts be fully functional in the target context of reception if initiators cannot always provide crucial information?

A quick review of the literature published by the Localisation Industry clearly shows that the objective of the initiators for localised texts is for them to be received as if they had been originally produced in the target locale. According to the Localisation Industry Standard Association (LISA), the goal of the localisation process is to release texts "with the look and feel of locally made products" (LISA 2003:5). This same goal is phrased somewhat differently in a later edition of the same publication:

> Localisation needs to go beyond language questions to address issues of content and look and feel with the ultimate aim of releasing a product that looks like it has been developed in country (LISA, 2004:11).

If viewed from a theoretical perspective, the objective of releasing texts that are received as original texts can be found in several proposals for translation typologies. This is the case of "instrumental equifunctional translations" by Nord (1997), "communicative translations" by Reiss and Vermeer (1984) and Newmark (1988), or "covert translations" by House (1997). Even when the theoretical backgrounds behind these proposals are clearly different, all researchers agree that the role of the target text is to be received as an original production in the linguistic and cultural region of reception. In the case of localisation, it should be mentioned that all texts that are translated, such as web text, software interfaces, videogames or small device interfaces, share this same goal. However, there is a lack of descriptive studies that describe the specific features of spontaneously produced digital texts in the context of reception that need to be incorporated into target texts. One of the questions that this paper can answer is whether the Industry has researched the specific conventional features in each locale, as the current approach is to review the translated product in-country or by a native editor. The need to identify conventions had also led some companies to introduce some innovative approaches, such as "crowdsourcing" (O'Hagan 2009), a process by which the community of web users propose possible translations and vote on them.[1]

In order to contextualise this study it should be noted that, according to a vast number of studies, translated texts show certain features that are inexistent in spontaneously produced texts (Baker 1995; Olohan and Baker 2001; Kenny 2001; Laviosa 1997; Tirkkonen-Condit 2004; etc.). The differentiated nature of translated language has led several scholars to coin different terms for this concept, such as "the third code" (Frawley 1984), "the third language" (Duff 1981) or "hybrid language" (Trosborg 2000). This is due to the fact that translation is a "a communicative event which is shaped by its own goals, pressures and context of production" (Baker 1996:175). Any translation process is subject to a number of constraints that are not equally present during the production of original texts, such as social, cultural, ideological and cognitive constraints (Baker 1999:285). These specific pressures will have an effect on the textual configuration of localised texts that can be identified using corpus linguistic techniques, both through parallel or comparable corpus studies. Additionally, previous research by the author has shown that technological constraints constitute an additional variable in this specific process (Jiménez-Crespo 2008a). For example, one of the most significant constraints that the translator has to face is the 'deconstruction' of the text to localise (Jiménez-Crespo 2008a; Dunne 2006). Thus, it is customary for localisers to work with texts that have been previously pre-segmented and, often, they are requested only to translate whichever segments do not have a match in existing TM databases. This technological environment places an undue burden on translators since

the cohesion and coherence mechanisms necessary in all texts (de Beaudegrande & Dressler 1981), might be somewhat absent during the comprehension stage of the translation process. Additionally, some features that are culture-dependent, such as text structures, are usually fixed (Neubert & Shreve 1992), and any changes to most parts of the textual structure would require extensive engineering at the Internationalisation stage. Thus, localisers are usually not allowed or discouraged from making changes to a fixed textual structure.

This point leads to the main criticism that can be directed towards the objective of the industry for localised texts: in the light of over a decade of empirical quantitative corpus-based translation research (Laviosa 2002), is it still feasible to defend the claim that a localised text will be similar to an ideal text that "looks like it has been developed in country" (LISA, 2004:11)? If corpus-based Translation Studies findings are extended to the localisation process, the working hypothesis for this study is that localised texts will not necessarily be similar to spontaneously produced texts in any locale. So far, no empirical research has been carried out about the properties of localised texts that will distance them from original productions. In order to have a contrastive base, this paper asserts that conventional linguistic features are one of the most salient aspects that can be quantitatively measured and contrasted between original and localised texts. Conventions play an important role in interactive digital texts (Nielsen & Tahir 2002) and in principle, translators should replace any conventional feature of the source texts in the target text with the established conventions in any given textual genre in the sociocultural context of reception (Nord 1997). Because of the significant role that this notion plays in this paper, the notion of 'convention' will be defined in the next section. Its implications for the localisation process will also be explored from a theoretical and practical perspective.

2.1. The notion of convention

As mentioned above, conventions are culture-dependent. This implies that they can potentially differ between similar genres in different cultures (Nord 1997: 54). As such, they play an important role during any translation process. Translators need to be aware of any conventional feature in the source text and be capable of replacing this with the established convention in the target culture. Conventions are also crucial during any communicative process, original production or any reading or comprehension process. From a philosophical perspective, the notion of convention has been defined as regularities in human behaviour in situations of co-operation (Lewis 1969). In Translation Studies, the notion of convention has been defined as:

> Implicit or tacit non-binding regulation of behaviour, based on common knowledge and the expectations of what others expect you to expect from them (etc.) to do in a certain situation (Nord 1991: 96).

From this definition it should be stressed that conventions are non-binding, an aspect that distances this notion from that of "norm" (Toury 1995).[2] In digital texts, non-compliance with a convention, such as placing the navigation menu to the right of the screen as opposed to the left, might slow down the communication process, but it will not stop it. On the other hand, non-compliance with a norm, such as including recurring spelling errors, will produce a negative effect on the user that will associate it with a lack of quality on the part of the company that released the text (Jeney 2007). This could lead to a lack of credibility that would stop the text from achieving its pragmatic goal.

The concept of convention is widely used in most theories and paradigms in Translation Studies. It is usually associated with different levels, such as "genre conventions" (Reiss and Vermeer 1984), "style conventions", "conventions of non-verbal conduct"[3] or "translation conventions" (Nord 2003; 1997). Most publications with an Industry perspective also associate this notion to different aspects, such as "cultural, language, business conventions" (LISA, 2007). In software localisation, Microsoft lists conventional aspects that are associated to the notion of locale:

> The locale determines conventions such as sort order; keyboard layout; and date, time, number, and currency formats. In Windows, locales usually provide more information about cultural conventions than about languages (Microsoft 2003).

In this case, it should be wrong to assume that language conventions are less important than cultural conventions. In fact, the dominating position of a limited number of software products that were localised helped conventionalise their linguistic features in most languages. This was done through a localisation process from English into other languages. Nevertheless, the number of web texts directly produced in most languages is nowadays higher than the number of software products. Thus, it is likely that the emergence and evolution of culture-dependent conventions will be more active in web texts.

Despite the fact the notion of 'convention' clearly differs between Translation Studies and the Localisation Industry, this brief review has highlighted that this concept frequently appears in studies and publications from both perspectives. Additionally, this critical review shows that certain theoretical and practical implications are not fully applied to the practice of professional localisation. The following is a summary of some basic theoretical aspects that can help clarify its essential role:

a) Alternatives and variants to conventions. Without an alternative, a convention as such cannot exist (Göpferich 1995). In this sense, any convention requires "alternatives" and "variants". An alternative consists of a linguistic form that is not conventional but that can accomplish the same communicative goal. For example, it is conventional in web pages to

include a link with the lexical unit "contact us" (Nielsen & Tahir 2002). A possible alternative could be "get in touch with us", "how to reach us" or "call or email us". Nevertheless, the former lexical unit is present in 89% of all commercial websites in English (ibid), even when a wide array of lexical units could accomplish this same function. A variant is the reduced array of variation that is accepted in any given convention. As an example, in Spanish corporate websites the most used lexical units in this case would be *contacto, contactar, contáctenos, información adicional* or *contacte con nosotros* (Jiménez-Crespo 2008b). Any translator has to be acquainted not only with the most conventional feature in any given genre, but also with the possible variants. In fact, in most non-technical genres stylistic variation might even be required (Gamero 2001: 54).

b) Conventions are arbitrary. The process by which conventions come into existence is totally arbitrary, provided that all possible alternatives can successfully accomplish the same communicative goal (Lewis 1969: 70). The fact that they are established at random means that they can differ from culture to culture (Gläser 1990: 29). Thus, this paper asserts that there is a need for contrastive corpus studies of the most localised genres in order to quantitatively identify which linguistic features can be considered conventional. In this regard, several scholars have pointed out that translators' intuition might not provide a valid judgment in most cases (Nord 2003; Hurtado Albir 2001). As this paper will show, most translators would assume that the most conventional translation for the lexical unit "about us" in Spanish would be *acerca de* or *acerca de nosotros*. This is the existing translation in two of the most used localised websites in Spanish contexts according to Nielsen Ratings, Google and Microsoft. Nevertheless, a previous descriptive study by the author (Jiménez-Crespo 2008b) showed that in the same digital genre, the most conventional lexical units are *La empresa* 'the company' or *¿Quiénes somos?* 'Who are we?'

c) Active and passive competence of genre conventions. The crucial role of conventions in professional activity and translator training is due to the distinction between active and passive competence (Gläser 1990: 72). Active competence can be defined as the ability of speakers of a language to recognise and produce the conventional features of textual genres, such as writing a resume or an email. Nevertheless, most speakers might not be able to produce certain textual genres, such as a patent, a purchase contract or a privacy policy on a website, even when they might recognise prototypical instances of the genres and be able to identify the possible range of variation. This is referred to as passive competence (Gamero 2001: 53). In localisation, most members of a discourse community have an active competence in writing emails or blogs, but the ability to produce a privacy policy or an effective homepage has to be consciously developed. Learning what is conventional in any specialised genre requires a systematic and conscious effort that allows the speaker

consciously to develop an active competence (Gläser 1990: 27). The lack of active competence on any given textual genre has been referred to as "genre deficit" or "text type deficit" (Hatim & Mason 1997: 133):

> Within a given language and across languages, the various forms of a given type may not be equally available to all users – a factor we may refer to as **text type deficit** (Hatim & Mason 1997: 133).

d) Evolution and flexibility. Genre conventions are not totally stable throughout time, but on the contrary, they evolve and change. Therefore translators need to be aware of this possible evolution both in time and space (Göpferich 1995). The evolution throughout time is of special interest to this study owing to rapid development in all technological fields. This evolution can be due to changes in a given culture, changes in the co-ordination problem that gave rise to a specific convention (Lewis 1969), and finally, certain interferences due to borrowings or mistranslations might be gradually assimilated and accepted as valid and correct. Many examples of this last instance can be found in digital texts. According to Bouffard and Craignon (2006), even when *contactez-nous* in French or *contáctenos* in Spanish could be considered borrowings, most speakers would consider these lexical units as valid choices in a website.

e) The role of conventions in localisation and translation. Genre conventions play an important role in the identification and translation of most technical and localised genres (Nord 1997: 53). First of all, they function as signs that facilitate the recognition of a given genre. Secondly, they activate the expectations of the reader. And finally, they are signs that co-ordinate the text comprehension process (Reiss & Vermeer 1984: 189). Therefore, given that translation entails both a textual comprehension and a textual production process, conventions also play a crucial role in it (Göpferich 1995: 168; Nord 1997, 2003). It should be noted that from a functionalist perspective, the substitution or adaptation of the conventions in the source text for those in the sociocultural context of reception is not automatic, but depends on the *skopos* of the translation and the norms of the target culture (Reiss & Vermeer 1984: 194). Nevertheless, it is logical to assert that most localised genres need to be functional texts in the target culture, and in principle all target texts should incorporate whichever conventions are established in the receiving locale.

f) Conventions in web design and localisation. The facilitating role of conventions in web usability has been recognised by most researchers in the field (Nielsen & Loranger 2007; Nielsen & Tahir 2002; Brink *et al.* 2002). The interactive nature of digital texts, the space constraints of screen displays or the slower pace at which receivers process screen texts imply that they need to be extremely clear and concise. They also need to incorporate the established conventions in each digital genre. The

importance of following the conventions in digital texts was pointed out by Nielsen & Tahir (2002: 37).

> […] over time, we expect more and more conventions to emerge.[...] by the time a user arrives at your homepage for the first time, that user will already be carrying a large load of mental baggage, accumulated from prior visits to thousands of other homepages. […] by this time, users have accumulated a **generic mental model** of the way homepages are supposed to work, based on their experiences on these other sites.

This "generic mental model" consists of the set of conventions shared by a specific discourse community for a specific genre, such as corporate websites or blogs. This model does not only include linguistic and cultural features, but it also includes typographical, graphical or functionality aspects. As an example, any web user has a generic mental model of what happens when a word is typed in a search box and the 'search,' 'ok' or 'go' button is activated.[4] This conventional search result page is part of the genre prototype expected from most websites. An empirical study by Vaughan & Dillon (2006) concluded that digital texts that follow structural and semantic conventions provide higher usability, comprehension, performance and more effective navigation. Nevertheless, it should be pointed out that in order for convention to exist, there has to be a number of other alternatives that can accomplish the same communicative goal (Göpferich 1995). Therefore the authors also noted that usability levels in non-conventional texts improved gradually after each consecutive use. The experimental variables in this study included the placement and terminology used in navigation menus. In fact, the interactive and recurring nature of navigation menus means that these textual segments are highly conventional. This aspect has been researched in different languages such as English (Nielsen & Tahir 2002), French (Bouffard & Craignon 2006) and Spanish (Jiménez-Crespo 2008a, 2008b). Consequently, terminology used in navigation menus constitutes an optimal point in order to research whether localised texts follow established conventions in digital genres. Additionally, in previous research the author has shown that navigation menus in Spanish corporate websites are highly conventional (ibid).

Navigation menus have a metatextual function in any website. They are a visual representation of the superstructure of any website, and they facilitate the comprehension process by situating each page in the user's global mental model of the global website. According to Price and Price (2002: 70-73), the function of these textual segments is to facilitate the interaction between the user and the website, being part of the website's 'interface texts.' The rest of the texts are referred to as 'content texts,' that is, all the specific contents from each page or hypertextual node. Additionally, in large websites navigation menus are usually presented hierarchically in order to reduce a user's cognitive load (Spyridakis 2000).

This review of the notion of convention and its implications for localisation has shown that it plays a significant role that needs to be better understood through descriptive and contrastive corpus studies (Hurtado Albir 2001). The next logical step is to observe whether localised and original texts share the same conventional features or whether, on the contrary, genre conventions in source digital texts are assimilated in the target texts, a recurring issue in most translated texts (Reiss & Vermeer 1984). This effect has also been referred to as "interference" (Toury 1995).

3. Empirical study: Methodology

Corpus linguistic approaches to the study of translated texts have been extremely productive during the last couple of decades. These approaches have benefited immensely with the introduction of electronic corpora. A carefully constructed electronic corpus can be a source of conceptual, terminological and linguistic information as well as a quantitative base in order to research translation processes and products (Olohan 2004). It can be defined as a large principled collection of machine-readable texts that has been compiled according to a specific set of criteria with the goal of being representative of a target textual population. Among different corpus types (Laviosa 2002: 34-38), a comparable corpus includes a section of spontaneously produced texts and a second section of translated texts. According to Mona Baker (1995: 234) a comparable corpus can be defined as:

> Two separate collections of texts in the same language: one corpus consists of original texts in the language in question and the other consists of translation in that language from a given source language or languages [...]. Both corpora should cover a similar domain, variety of language and time span, and be of comparable length.

As navigation menus were selected due to their highly conventional nature (Nielsen & Tahir 2002), a comparable subcorpus of navigation menus was extracted for the Spanish Comparable Web corpus compiled by the author (Jiménez-Crespo 2008a). This subcorpus contains exclusively lexical units from homepage navigation menus, the most comprehensive menus in any website. Given that text in some navigation menus is embedded in graphics, all lexical units where extracted manually and normalised to .txt format. The wider Spanish Comparable Web corpus was compiled as part of a broader project that has as a goal to research the process of localisation and its specific restrictions (*ibid*). This comparable corpus comprises a section of non-translated Spanish corporate websites (172 sites) and a section that includes all localised websites into Castilian Spanish from the largest North American corporations according to the Forbes list (95 sites). Both sections of the corpus were downloaded synchronically in one day during 2006. Corporate websites were selected as a representative genre in localisation because they are the most

conventionalised among all digital genres (Kennedy & Shepherd 2005). A detailed description of characteristics and specifications of the corpus can be found in previous publications by the author (Jiménez-Crespo 2008a, 2008b).

Spanish Evaluation Corpus	Original Section	Localised Section
Words total	1,815	3,558
Navigation Menus	172	95
Words/navigation menu	10.55	37.45
Types	407	935
Average word length	6.59	6.42

Table 1. Description of Spanish comparable navigation subcorpus.

The Comparable Web Navigation Corpus comprises 5,373 words. The Spanish non-translated section of the corpus includes 1,845 words, an average of 10.45 words per navigation menu. The localised section of the corpus includes 3,558 words, with 37.45 words per navigation menu. In order to control the possible effect of lexical repetition in web corpora (Jiménez-Crespo & Tercedor 2008), only one navigation menu was selected from each website. The quantitative differences among both sections of the corpus are due to the presence of large multinational corporate websites in the localised corpus, while the original corpus includes websites from large, medium and small Spanish corporations. Nevertheless, as the goal of this paper is to study conventions shared by a specific discourse community, namely all Spanish web users, the presence of both large and small Spanish corporate sites the original corpus is justified.[5] Thus, original navigation menus belong on average to smaller sites and as a consequence, the superstructure represented by lexical units in the website menu is more limited. Additionally, this methodological decision was due to the fact that only large US international corporations offer a localised website for Spain, and not simply an 'international Spanish' site.[6]

The next step was to assign each lexical unit to the genre prototypical superstructure obtained through a previous descriptive study (Jiménez-Crespo 2008b). The main communicative blocks present in corporate websites are (a) homepage, (b) contact us, (c) about us, (d) product-services, (e) news, (f) legal, (g) user's areas and (h) interaction. In a second level, each communicative block includes different communicative sections: sixteen different sections were identified in the block "about us", such as "location", "experience", "quality" or "history". Each lexical unit compiled in the corpus represents a concept associated to the prototypical genre's superstructure. The following graphic shows a visual representation of this superstructure and the different concepts to which each lexical unit was assigned.
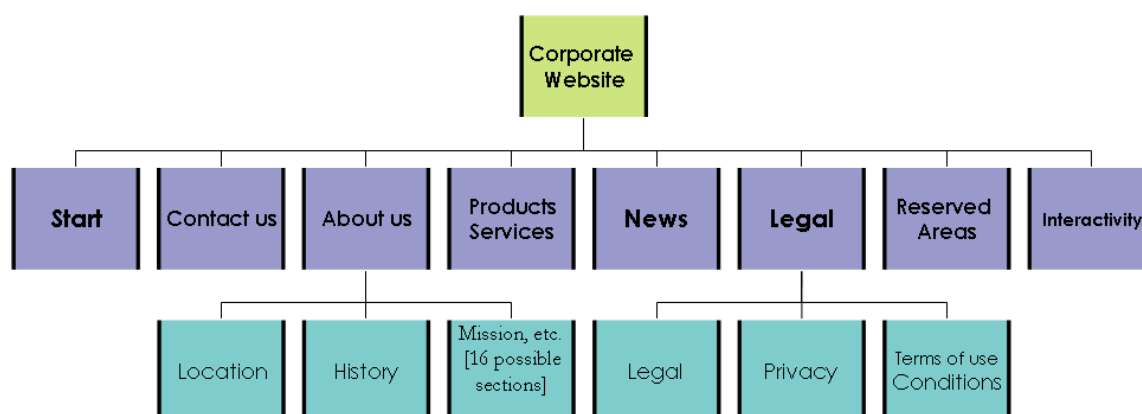
Figure 1. Corporate website prototypical superstructure.

Each lexical unit was identified with the concept in the structure of the text that it represents. This will allow contrasting terms used for the same concept in original and localised texts, and therefore it will make it possible to observe any differences in the conventions used. As an example, most corporate websites include a communicative section called "mission", this genre-based analysis made it possible to observe that in Spanish this concept can be represented by the following lexical units: *misión, filosofía, compromiso social, enfoque, fundamentos, visión* or *trabajamos para ti.*

4. Results

The first concept that was contrasted was the most conventionalised lexical unit in English texts, 'contact us.' Its selection was due to the fact that this lexical unit is highly conventionalised in English web texts. According to Nielsen & Tahir (2002), it appears in 89% of US corporate websites. The possible variation in the expression of this concept has also been previously explored in French (Bouffard & Craignon 2006). As seen in Figure 2, the study identified 23 lexical units associated to this concept in localised texts, while only 16 possible alternatives were identified in original Spanish texts. From all localised forms, 12 are not present in original Spanish texts and therefore could be identified as possible alternatives to the conventions. The combined frequency of these alternative units is 24.73%. Even when they might be correct syntactically or lexically, they could be considered "unique items" (Tirkkonen-Condit 2004), items that are present only in translated texts.

In Spanish corporate websites, the most frequent term associated with this concept is the noun *contacto*, 49.66%, followed by the verb *contactar* 14.48%. These could be identified as the conventional lexical units in original texts. Nevertheless, the most frequent form in localised texts is *contáctenos*, 21.11%, followed by *contacte con nosotros*, 13.33%. These lexical units are direct transfers of the conventional form in English source

texts. The influence that this convention exerts on the localisation process can be observed through the use of the pronoun *nosotros*, 'us´. In original texts, only 15.86% of all identified units include this pronoun, either linked syntactically with the Spanish preposition *con* or attached enclitically, '-nos´, while 51.11% of all localised forms incorporate this pronoun. In fact, all lexical units that incorporate this pronoun show higher frequencies in localised texts than in original ones. These forms have been identified with a box in Figure 2. Additionally, this influence can be observed through a contrastive analysis of the use of verbs in these lexical units. The source English convention 'contact us' is formed with a verb, while the most conventional lexical unit in Spanish texts is formed with a noun. In original texts, a verb is included in 41.38% of all lexical units, while 71.11% of localised ones include a verb. This clearly illustrates how localised texts are influenced by a conventional feature of source texts and therefore established genre conventions in the target culture are not necessarily followed.
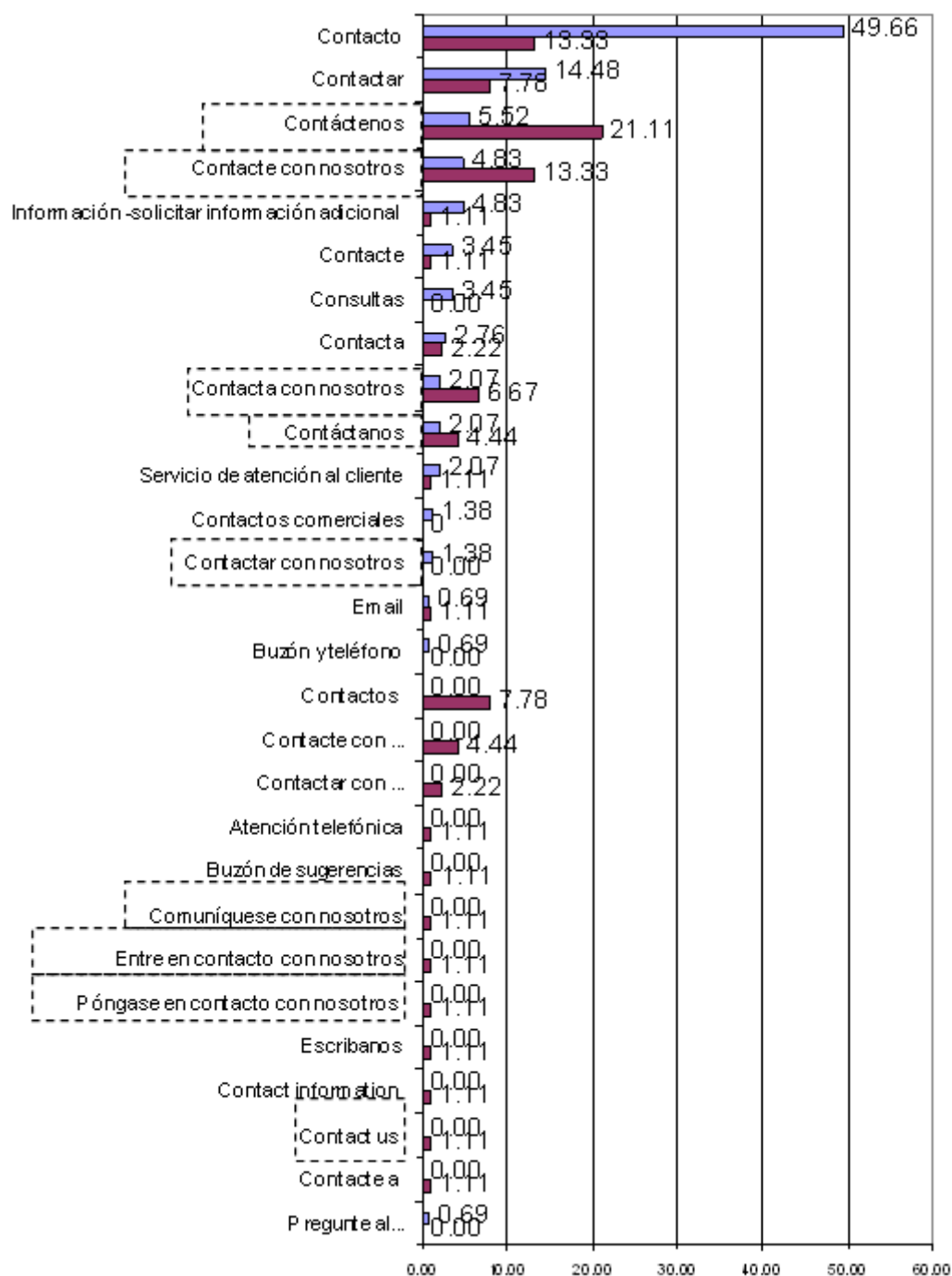
Figure 2. Contrastive table of original and localised lexical units for the communicative block 'contact us.'

The second most frequent communicative block in commercial websites is "about us". It includes information about the company, such as its locations, history, experience, mission or publications. The concept associated to this communicative block was selected because it is conventional to use the lexical unit "about us" in 71% of English corporate websites (Nielsen & Tahir 2002). In Spanish websites, the most conventional lexical units for this concept are *empresa* or *compañía*, 'the company', 37.2%, followed by *¿quiénes somos?*, 'who are we?' and *presentación*, 'presentation'.
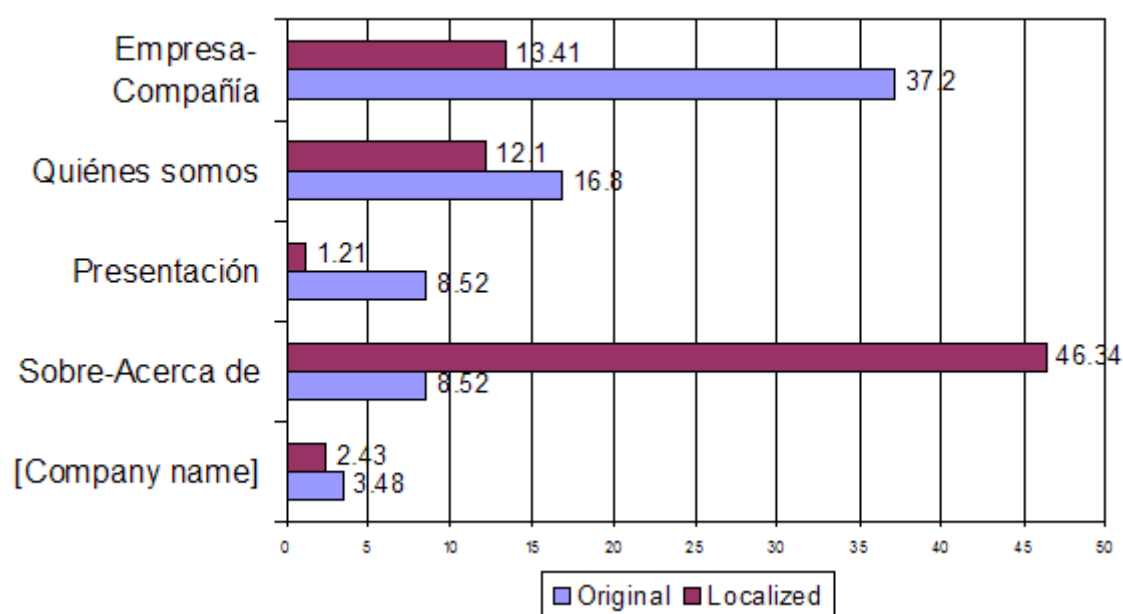
Figure 3. Contrastive study of original and localised lexical units for the communicative block 'about us.'

As Figure 3 shows, in this case localised sites do not comply with conventions in original websites. Localised websites are highly influenced by the conventional 'about us' in source texts. In Spanish, the preposition 'about' corresponds to two different synonynimical prepositions, *sobre* and *acerca*. The combined frequency of use of these two prepositions in localised navigation menus is 46.34%, while they represent only 8.52% of use in original websites. Additionally, in original Spanish texts the frequency of the preposition *sobre* is higher than that of the preposition *acerca*, even when this tendency is reversed in translated websites. This could indicate that not only target texts are influenced by conventional features in source texts, but that they are also directly transferred given that *acerca* resembles 'about' more so than the synonym *sobre.* This could indicate a "lexical priming" effect (Hoey 2005) given that both words start with the letter "a".[7]

This analysis has shown that influence of conventional linguistic forms in source texts in the most frequent superstructural concepts in websites can be observed in translated texts. In order to observe whether this effect also affects less common concepts, a final summative contrastive analysis of other superstructural concepts was carried out (Jiménez-Crespo 2008b). The following contrastive table shows the most conventional lexical units used in the identification of the most frequent communicative blocks and sections. The first column indicates the metatextual concept in the website's superstructure that the lexical unit represents. As the goal of the paper is to study conventional features, only the first and second most used lexical units are included in the table. This makes it possible to

contrast directly whether localised texts share the same conventions for navigation terminology as original texts.

Communicative block/section	Original Spanish Corpus		Localised Spanish corpus	
	Lexical Units	**Frequency**	**Lexical Unit**	**Frequency**
A.Start Page	Inicio	60%	Inicio	44.64%
	Home	30.36%	Home	24.04%
B. Contact us	Contacto	49.42%	Contáctenos	21.11%
	Contactar	14.18%	Contacte con nosotros	13.33%
C. About us	Empresa	37.2%	Acerca de…	30.48%
	Quienes somos	16.8%	Empresa	13.04%
1)Location	Localización	20.41%	Oficinas	21.82%
	Oficinas	20.41%	Donde comprar	10.91%
2)Company Experience	Clientes	36.66%	Casos de éxito	18.51%
	Trabajos-realizados	30%	Premios	18.51%
3)Mission	Misión	27.07%	Responsabilidad	21.05%
	Filosofía	10.6%	Compromiso social	18.42%
4) Quality	Calidad	37.5%	Calidad	45.45%
	Acreditaciones	12.5%	Certificaciones	18.18%
5)History	Historia	100%	Historia	84.84%
			Historial	3.03%
D. News/Press Releases	Noticias	40.63%	Noticias	34.11%
	Novedades	14.58%	Sala de prensa	17.02%
	Notas de prensa	9.57%	Novedades	9.57%
3)Events	Eventos	100%	Eventos	83.33%
			Actos	8.33%
E. Products -Services				
1)Products	Productos	68.47%	Productos	56.25%
	Catálogo	19.56%	[Produtct name]	35%
2)Services	Servicios	88.31%	Servicios	53.97%
	Soluciones	49.6%	Soluciones	25.40%
3) Promotions	Oferta	70.58%	Promoción	50.%
	Promoción	23.52%	Oferta	28.57%
F. Legal				
1) Legal information	Aviso legal	69.75%	Aviso legal	22.73%
	Nota legal	6.82%	Información legal	15.91%
2)Privacy Policy	Política de privacidad	52.17%	Política de privacidad	33.80%
	Privacidad	21.74%	Privacidad	26.76%
3)Terms and Conditions	Condiciones de uso	37.50%	Condiciones de uso	29.73%
	Normas de uso/ Condiciones generales/Términos de uso/Términos y condiciones	12.50%	Términos y condiciones	24.32%
G. User's Areas	Acceso clientes	24.32%	Log in	10.81%
	Clientes	10.81%	Clientes	8.10%
1) Careers	Bolsa de empleo	12.50%	Empleo	14.55%
	Empleo	8.33%	Carreras	12.73%
8) Investors	Inversores	33.23%	Inversores	20.59%
	Información financiera/ Información para accionistas/Relaciones con los Inversores /Investor relations	16.67%	Relaciones con los inversores/ Investor Relations	14.71%
H. Interactivity				
1) Site map	Mapa web	54.21%	Mapa del sitio	41.64%
	Mapa del web	14.3%	Site map	18.33%

2) Search	Buscar	69.23%	Buscar	54.10%
	Buscador	11.54%	Búsqueda	21.13%
3) FAQ	Ayuda/ FAQ	27.78%	Ayuda	24.24%
	Atención al cliente	16.67%	Preguntas frecuentes	18.18%
4) Links	Enlaces	50%	Enlaces	33.33%
	Links	35%	Enlaces rápidos	20%
5)Register	Registrese	14.29%	Registro	19.05%
	Hagase cliente Registro	13.33%	Subscribase Regístrese	14.29%

Table 2. Contrastive table of conventional original and translated terminology in corporate websites.

As a last step, the average frequencies of the most used lexical units were calculated for both sections. In original texts, the average frequency of the most used lexical unit, such as *contacto* or *empresa*, was 58.26%. Therefore it could be deduced that the terminology used in original Spanish websites is conventional since its average frequency is higher than 50% (Gamero 2001). However, the average frequency of the first lexical unit in all communicative blocks in localised texts is 43.39%, significantly lower than original texts and below the threshold to be considered conventional (Nielsen 2004; Gamero 2001; Gläser 1990). Consequently, it could be indicated that the terminology used in localised texts shows fewer conventionalisation levels and higher denominative variation. To some extent, this is detrimental to the crucial goal of usability (Nielsen & Loranger 2006; Vaughan & Dillon 2006). Less conventional navigation terminology means that users have to make a greater conscious effort to interpret the lexical units used.

	Original Section Average	Localised Section Average
Most frequent lexical unit	58.26%	43.39%
2nd most frequent lexical unit	19.05%	13.26%
Frequency of 1st and 2nd combined	77.31%	56.65%
Frequency of the most conventional lexical unit in original texts	58.26%	34.36%
Percentage of errors		6.82%

Table 3. Summary. Average frequency of conventional lexical forms in both sections of the corpus.

As mentioned previously, any convention needs accepted alternatives that can accomplish the same communicative function, the so-called variants. If the average frequency of a second lexical unit is factored in, the average frequency in original texts is 77.31%. In localised texts this combined average frequency is 56.25%, approximately 20% lower.

However, the most relevant calculation in the study of convention would be to observe the average frequency in localised texts of the established conventions in original Spanish sites. This could be indicative of whether localised navigation menus comply with the established conventions on original sites, such as the use of *contacto* or *empresa*. In this case, the average frequency of the most conventional lexical unit in original texts is 58.26%. In localised texts, the frequency of these same lexical units drops to 34.36%, significantly lower than the average of original texts and also lower than the average frequency of the most used lexical form. The difference of 24.39% clearly indicates that localised texts do not necessarily follow the established conventions in original textual genres in a given discourse community. As observed previously, this is due to two differentiated factors. On the one hand, localised texts show the recurrent and common interference of the conventions of source texts (Reiss and Vermeer 1984: 194). On the other hand, this difference is indicative of the different constraints during the localisation-translation process (Baker 1999), such as the decontextualisation of the textual segments, the translation by a group of localisers or the effect of Translation Memory use, that lead to higher levels of denominative variation in the final product (Jiménez-Crespo 2008a).

In part, this effect could be attributed to the relative lack of use of reference corpora in this field, as well as the lack of descriptive studies in the most localised genres in different locales. Thus during the localisation process, translators and editors might not have available genre-based quantitative studies or representative quantitative data that can assist them in making judgments or decisions about (1) what can be considered a conventional feature in a source text and (2) which conventional feature in the target text accomplishes the same communicative goal. This study has already been completed in the English-Spanish combination and in the most translated digital genre, the corporate website (ibid; Jiménez-Crespo 2008b). These types of studies have been produced for many legal, scientific and technical genres, and they are crucial in localisation given that the foundation of web usability consists in complying with any established conventions.

Finally, the contrastive analysis of the terminology used in navigation menus made it possible to identify the percentage of errors in these recurring textual segments, 6.82%. As an example, in Figure 3 the lexical unit *contactos* 'contacts´ was only used in localised texts. The plural form of this noun is associated in Spanish to the concept of 'personal ads´ or 'personal profiles', etc. Navigation menus need to be devoid of any ambiguity in order to be effective (Jeney 2007; Price & Price 2002). Therefore even when in principle this would be correct, it could be considered as a translation error, since Spanish users would firstly associate this term with a different concept.

5. Conclusions

The aim of this article was twofold: on the one hand it reviewed from a theoretical perspective the role of convention in translation, and more specifically in localisation, and on the other hand it researched empirically whether localised texts are fully functional texts in the target context of reception. This aspect was associated with the replacement of conventional features of source texts by those established in target genres. From a functionalist perspective, it was argued that localised texts should be adapted to include conventionalised linguistic features in the target textual population. Thus this adaptation entails producing texts that comply with the recommendations from usability publications (Vaughan & Dillon 2006; Nielsen & Loranger 2007; Brink *et al.* 2002).

The first conclusion from this study is that the role of conventions is crucial in both web and software localisation. It has been shown that conventions are culture-dependent and do not automatically transfer from one culture to another. In this sense, conventional features can only be deduced from wide corpus-based studies (Nord 2003; Hurtado Albir 2001). This study points out the need for empirical corpus-based studies in order to produce both (1) descriptive studies that can be used as a translation tool and for evaluation purposes, and (2) contrastive studies that can observe which conventional features of source texts are frequently transferred directly during the translation process.

The results of the empirical corpus study have clearly demonstrated that localised texts do not follow the conventions in original websites (Jiménez-Crespo 2008b), as they are heavily influenced by conventional features of original US source texts. Therefore even when the main goal of the Localisation Industry is releasing localised texts that are received as locally-made products (LISA 2003:5), it has been observed that distinctive constraints during the localisation process lead to target texts with distinctive linguistic features. The findings are coherent with previous corpus studies that have identified distinctive features in translated texts in different language combinations (Baker & Olohan 2001; Laviosa 1998; Kenny 2001; Tirkkonen-Condit 2004). From a practical perspective, the implications of this study can be helpful for practitioners and editors in achieving fully functional target texts that are received as originals.

However and as a final remark, these results cannot absolutely answer whether localised sites are received or function as "locally-made" websites. This paper does not prove that carefully edited localised texts that do not follow established conventions are less usable than those that do comply with them. As mentioned above, this is a guiding principle in web usability, but empirical studies such as those carried out by Vaughan

& Dillon (2006) would be needed to find out what kind of effect is brought about by the direct transfer of source text conventions from a hegemonic culture and language. In a young translation modality such as localisation, the possibilities for future research are endless.

Bibliography

- **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 Honor of Juan C. Sager*. Amsterdam/Philadelphia: John Benjamin's, 175-186.

- **—** (1999). "The Role of Corpora in Investigating the Linguistic Behavior of Professional Translators." *International Journal of Corpus* Linguistics 4 (2), 281-298.

- **—** (1993). "Corpus Linguistics and Translation Studies: Implications and Applications." Mona Baker, Francis Gill & Elena Tognini-Bonelli (eds) (1993). *Text and Technology: In Honour of John Sinclair*. Amsterdam/Philadelphia: John Benjamins, 233-250.

- **de Beaudegrande**, Robert & Wolfgang Dressler (1981). *Introduction to Text Linguistics*. London: Longman.

- **Bouffard, Paula** & Phillippe Caignon (2006). "Localisation et variation linguistique. Vers une géolinguistique de l'espace virtuel francophone." *Meta* 51 (4), 806-823.

- **Brinck, Tom**, Gergle, Darren & Scott D. Wood (2002). *Usability for the Web*. San Francisco: Morgan Kauffman.

- **Cronin, Michael** (2003). *Translation and Globalisation*. London: Routledge.

- **Dunne, Keiran** (2006a). "A Copernican Revolution." In Keiran Dunne (ed.) (2006). *Perspectives on Localisation*. Amsterdam: John Benjamins, 1-11.

- **—** (2006b). "Putting the cart behind the horse: rethinking localisation quality management." Keiran Dunne (ed.) (2006). *Perspectives on Localisation*. Amsterdam: John Benjamins, 95-117.

- **Duff, Alan** (1981). *The Third Language: recurrent problems of translation into English*. Oxford: Pergamon Press.

- **Esselink, Bert** (2006). "The evolution of localisation." Anthony Pym, Perekstenko, Alexander & Bram Starink (eds) (2006). *Translation Technology and its Teaching.* Tarragona: Intercultural Studies Group, 21-30.

- **Folaron, Debbie** (2006). "A discipline coming of age in the digital age." Keiran Dunne (ed.) (2006). *Perspectives on localisation*. Amsterdam/Philadelphia: John Benjamins, 195-222.

- **Frawley, William** (1984). "Prolegomenon to a theory of translation." William Frawley (ed.)(1984). *Translation. Literary, linguistic and philosophical perspectives*. London/Toronto: Associated University Presses, 159-175

- **Gamero Pérez**, Silvia (2001). *La traducción de textos técnicos*. Barcelona: Ariel.

- **Gläser, Rosemarie** (1990). *Fachtextsorten im Englischen*.Tübingen: Narr.

- **Göpferich, Susanne** (1995). *Textsorten in Naturwissenschaften und Technik. Pragmatische Typologie- Kontrastierung-Translation*. Tübingen: Gunter Narr.

- **Gouadec, Daniel** (2003). "Le bagage spécifique du locateur/localisateur. Le vrai «nouvel profil» requis." *Meta* 28 (4), 526-545.

- **Hatim, Basil** & Ian Mason (1997). *The Translator as Communicator.* London: Routledge.

- **Hoey, Michael** (2005). *Lexical Priming. A new theory of words and language*. London and New York: Routledge.

- **Holz-Mänttäri, Justa** (1984). *Translatorisches Handeln. Theorie und Methode*. Helsinki: Suomalainen Tiedeakatemia.

- **House, Juliane** (2001). "Translation Quality Assessment: Linguistic Description versus Social Evaluation." *Meta* 46 (2), 243-257.

- **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, Miguel** A. (2008a). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. PhD thesis. University of Granada, Spain.

- **—** (2008b). "Caracterización del género 'sitio web corporativo' español: análisis descriptivo con fines traductológicos." Mariela Fernández Sánchez & Ricardo Muñoz Martín (eds) (2008). *Aproximaciones cognitivas al estudio de la traducción e interpretación*. Granada: Comares, 259-300.

- **—** & Maribel Tercedor (2008). "Web corpora in Translation Studies: Theoretical and methodological issues." Paper presented at *New Trends in Language Teaching and Translation Studies: International Conference on Corpus Linguistics* (University of Granada, 23-25 September 2008).

- **Kennedy, Alistair** & Michael Shepherd (2005). "Automatic Identification of Home Pages on the Web." *Conference Proceedings, XXXVIII Annual Hawaii International Conference on System Sciences, Maui, Hawaii*. Los Alamitos, CA: IEEE-Computer Society, on line at http://users.cs.dal.ca/~shepherd/pubs/home_pages.pdf, 1-9 (consulted on 05.06.2009).

- **Kenny, Dorothy** (2001). *Lexis and Creativity in Translation. A corpus-based study*. Manchester: St. Jerome.

- **Kussmaul, Paul** (1995). *Training the Translator*. Amsterdam/Philadelphia: John Benjamins.

- **Laviosa, Sara** (2002). *Corpus-based Translation Studies*. Amsterdam: Rodopi.

- **Lewis, David** K. (1969). *Convention. A Philosophical Study*. Cambridge, MA: Harvard University Press.

- **LISA** (2004). *Localisation Industry Primer*, 2nd Edition. Geneva: The Localisation Industry Standards Association (LISA).

- **—** (2003). *Localisation Industry Primer*. D. Fry (ed.). Geneva: The Localisation Industry Standards Association (LISA). Online at: http://www.lisa.org/products/primer.html (Consulted 05.05.2008).

- **Mazur, Iwona** (2007). "The metalanguage of localisation: Theory and practice." *Target* 19 (2) 2, 337-357.
- Microsoft Corporation (2003). *Developing International Software*. Redmond: Microsoft Corporation.

- **Neubert, Albrecht** & Gregory Shreve (1992). *Translation as Text*. Kent, Ohio: Kent State University.

- **Newmark, Peter** (1988). *A Textbook of Translation*. Oxford: Pergamon Press.

- **Nielsen, Jacob** & Hoa Loranger (2006). *Prioritizing Web Usability*. Indianapolis: News Riders.

- **—** (2004). "The need for web design standards." On line at: http://www.useit.com/alertbox/20040913.html (Consulted 18.07.2008).

- **—** & Marie Tahir (2002). *Homepage usability: 50 Websites deconstructed*. Indianapolis: News Riders.

- **Nord, Christiane** (1991). *Text Analysis in Translation: Theory, Methodolgy and Didactic Application of a Model for Translation-Oriented Text Analysis*. Amsterdam/Atlanta: Rodopi.

- **—** (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained*. Manchester: St. Jerome.

- **O'Hagan, Minako** (2009). "An exploration of the crowd as assessors of translation: translator training of popular entertainment in the context of crowdsourcing." Paper presented at *Monterey Forum 2009 Principles and Practices of Assessment in the Training of Translators and Interpreters* (Graduate Institute of Translation and Interpretation, Monterey CA, 2-4 April 2009)

- **—** & David Ashworth (2003). *Translation-Mediated Communication in a digital World: facing the challenges of Globalisation and Localisation*. Clevendon: Multilingual Matters.

- **Olohan, Maeve** (2004). *Introducing Corpora in Translation Studies.* London: Routledge.

- **—** & Mona Baker (2000). "Reporting that in Translated English. Evidence for Subconscious Processes of Explicitation?." *Across Languages and Cultures* 1(2), 141-158.

- **Price, Jonathan** & Lisa Price (2002). *Hot text. Web Writing that works*. Berkeley, CA: News Riders.

- **Pym, Anthony** (2003). "What localisation models can learn from Translation Theory." *The LISA Newsletter: Globalisation Insider* 12.

- **Reiss, Katherine** & Hans J. Vermeer (1984). *Grundlegung einer Allgemeinen Translationstheorie*. Tubinga: Niemeyer.

- **Schäler, Reinhard** (2002). "The Cultural Dimension in Software Localisation." Localisation Focus 1 (2).

- **Spyridakis, Jen** (2000). "Guidelines for Authoring Comprehensible Web Pages and Evaluating Their Success." *Technical Communication* 47 (3), 301-310.

- **Tirkkonen-Condit, Sonia** (2004). "Unique items – over – or under-represented in translated language" Anna Mauranen & Pekka Kujamäki (eds) (2004). *Translation Universals, Do They Exist?.* Amsterdam/Philadelphia: John Benjamins, 177-184.

- **Toury, Gideon** (1995). *Descriptive Translation Studies and beyond*. Amsterdam/Philadelphia: John Benjamins.

- **Trosborg, Anna** (2000). "Translating Hybrid Political Texts." A. Trosborg (ed.) (2000). *Analysing Professional Genres*. Amsterdam: John Benjamins, 145-158.

- **Vaughan, Misha** & Andrew Dillon (2006). "Why structure and genre matter for users of digital information: a longitudinal experiment with readers of a web-based newspaper." *International Journal of Human Computer Studies* 64: 502-525.

BIBLIOGRAPHICAL NOTES

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. He can be contacted at miguelji@rci.rutgers.edu.

---

[1] This approach has been taken by companies such as Facebook, Microsoft or Symantec (O'Hagan 2009).

[2] The concept of 'norm' can be defined as "the translation of general values or ideas shared by a group - as to what is conventionally right and wrong, adequate and inadequate - into performance instructions appropriate for and applicable to particular situations, specifying what is prescribed and forbidden, as well as what is tolerated and permitted in a certain behavioral dimension" (Toury 1998: 16).

[3] Such as icons, colors or typography.

[4] These are the three possible conventional alternative terms used in English websites according to Nielsen & Tahir (2002).

[5] This subcorpus was extracted from the Spanish Web Evaluation Corpus (Jiménez-Crespo 2008a), that has aproximately 20,000 original and 20,000 localised webpages.

[6] The reason behind this decision was to control an additional variable in the study, dialectal variation. As an example, Microsoft offers 19 different Spanish locales.

[7] Additionally, the preposition 'sobre' is more frequent in Spanish than "acerca" according to the Spanish corpus CREA (Reference Corpus of Contemporary Spanish). According to the word frequency obtained from the corpus CREA, "sobre" is the 32nd most frequent word in Spanish, while 'acerca' appears in the 758th position in the same wordlist.