The future of "universal" tendencies: a review of papers using localized websites

Miguel A Jiménez-Crespo

Rutgers University

Abstract: Since the emergence of Corpus-Based Translation Studies, research into potential regularities in translational behaviour or "general tendencies in translation" has been at its core. However, translation scholars have also criticized this type of descriptive research from methodological and theoretical grounds (i.e. Tymoczko 2005, 1998; Snell-Hornby 2006; House 2008). This paper reflects on these issues through an overview of studies on general tendencies using a translation modality, web localization. The goal is to discuss what this new modality as can add to the existing body of knowledge on general tendencies. Methodologically, the data for most studies was provided by the Spanish Web Comparable corpus containing 40,000 original and localized webpages from corporate websites. The overview of results will summarize the findings from published studies on explicitation, conventionalization and sanitation (--- 2008a, 2009a, 2009b, 2009c, forthcominga; Diéguez 2008), as well as an additional feature of translations, the tendency to "clone" source text structures (Larose 1998; --- 2009c, forthcomingb). This tendency has not been previously conceptualized as a general tendency, partly due to the fact that it cannot be directly studied using current methodologies and lexical analysis tools mostly focused on word-based metrics.

1. Introduction

The emergence of Corpus-Based Translation Studies in the 1990s was mostly focused on research into potential regularities in translational behaviour or 'general tendencies in translation' (Laviosa 2002). However, translations scholars have also criticized this type of descriptive research from methodological and theoretical grounds (i.e. Tymoczko 2005, 1998; Snell-Hornby 2006; House 2008). These have been mostly due to the tendency to overgeneralize the results obtained in limited translation subsets, such as literary or journalistic texts, as well as epistemological issues related to the use of the term 'universal' introduced by the seminal paper by Baker (1993). Nevertheless, the number of studies keeps increasing with adaptations and modifications to some of the most controversial issues, such as the 'universality' claim. A prototypical understanding of this cluster of tendencies has moved the focus away from a much maligned 'universal' notion, thus continuing research towards explaining and predicting existing phenomena (Chesterman 2010). The ongoing interest in this field of research can be witnessed by two international conferences mostly devoted to this topic in 2010, the "Methodological Advances in Corpus-Based Translation Studies" in College Ghent, Belgium, and "Using Corpora in Contrastive and Translation Studies" in Edge Hill University, UK. Following Chesterman´s call (2004: 47) to test these phenomena in different translation modalities and types, this paper offers an overview of recent corpus-based research using a newly established translation modality, web localization, as well as a relatively new digital genre, corporate website. This overview includes studies on explicitation in localised websites (Jiménez-Crespo forthcominga, 2010; Diéguez 2008) and conventionalization and sanitization (Jiménez-Crespo 2009a). Additionally, a tendency that does not rely in word-based metrics is proposed, 'cloned texts'. This notion was introduced by Larose (1998) in the context of quality evaluation in translation,

and it is related to the cloning of micro and macrostructures of source texts. It also involves a different methodology based on a genre-based approach (Swales 1990; Gamero 2001; Jiménez-Crespo 2008b). The proposal to conceptualize this already observed phenomenon as a general tendency is due to the renewed focus on structural cloning introduced by technology use in translation (Pym 2010; Bowker 2006), mostly related to the increasing use of translation memory tools (Jiménez-Crespo 2009b).

2. Web localization and the study of general tendencies

The technological revolution brought by the Internet has had a profound impact in the theory and practice of translation (Pym 2010; Munday 2008). The impact on the latter can be witnessed by the overall percentage of technology-related modalities in the translation industry, an estimated 40% in 2005 (Schäler 2005). As far as the impact in Translation Studies, an overview of current research trends shows that technology has prompted four new areas of inquiry: (1) empirical descriptive research into all types of localization from both professional and Translation Studies perspectives (e.g. Reineke 2005; Dunne 2006; Jiménez-Crespo 2008a), (2) Translation Technology tools (Austermhül 2000; Bowker 2002; Daelemans and Hoste 2009), (3) how to incorporate these emerging trends in translation training (Alcina 2008), and (4) ethical/philosophical implications of localization and globalization (Cronin 2003). This paper focuses on the first of these trends, applying the descriptive paradigm to the study of localisation.

The underlying rational to this review is that localisation offers the opportunity to review previously research in Descriptive Translation Studies (Holmes 1987; Toury 1995). More

3

specifically, this paper focuses specifically on the so-called translation 'universals' (Baker 1993), 'norms-laws' (Toury 1995, 2004), 'hypotheses' (Laviosa 1998), 'features' (Chesterman 2000), 'universal tendencies' (Mauranen 2008) or the term used in this paper, 'general tendencies' (Olohan 2004). This concept, despite ongoing conceptual and epistemological debates (Chesterman 2010), can be generally defined as "features which typically occur in translated text rather than original utterances and which are not the result of interference from specific linguistic systems" (Baker, 1993: 243). Normally, these features have been mostly researched using both parallel and comparable corpora (Baker 1995). The former have been instrumental in identifying 'S-universals', that is, hypotheses claiming "to capture universal differences between translations and their source texts, i.e. characteristics of the way in which translators process the source text" (Chesterman 2004: 39). Comparable corpora have been mostly related to the identification of 'T-universals', those related to the translation product, in which translated texts are compared to similar spontaneously-produced texts in the target language. To date, research into general tendencies using digital localised texts has only been done using a comparable corpus methodology, and therefore, focused on 'T-universals' tendencies. Despite an increasing number of parallel and comparable web corpora, to date the only comparable corpus that has been used to test these hypotheses is the Spanish Web Comparable Corpus (Jiménez-Crespo 2008a), while another interesting study was carried out by Diéguez (2008) in which a corpus that contained selected terminology found in US websites localised for the Chilean market was compiled.

2.1. Why web localisation?

Localisation is a novel translation modality that operates under specific procedural, communicative and cognitive constraints. For example, new translation technology in combination with content management systems, are forcing translators to process texts that are normally deconstructed (Shreve 2006) and it is not uncommon to work without a stable source text (Pym 2010). As far as the communicative context is concerned, most digital genres are highly interactive. They are read on screens, and this has lead usability scholars to prescribe specific writing styles to produce successful websites (Nielsen and Loranger 2006).These principles are based on cognitive efficiency, and they have been widely research through user-based empirical studies.

Most tendencies in translation were proposed and being tested prior to the appearance of the World Wide Web in the mid 90's, and consequently, prior to the emergence of localized web digital genres. In fact, Baker's (1993) seminal paper appeared several years before corporate websites were consolidated as a digital genre (Shepherd and Watters 1998). Many more digital genres, such as social networking sites, are still developing and evolving, and they are also leading to new translation models, such as crowdsourcing (O´Hagan 2009). If explicitation, sanitization or conventionalization can be considered general tendencies, they should also be confirmed in current or any future translation types and modalities, such as professional or crowdsourced web localization. In other words, any claim about general tendencies on translational behaviour should be tested in translations from seven year olds to those from the most reputed translator in the United Nations, or for that matter, localized websites. In any case, a prototypical understanding of these phenomena entails that it is perfectly legitimate to state that a general tendency occurs in specific subsets, such as non professionals or subtitling, but not in others (Chesterman 2010). This is one of the main benefits of leaving aside the much malign

concept of 'universal', and replacing it with 'tendencies' (Mauranen 2008; Olohan 2004) or 'features of translated language' (Chesterman 2004).

This modality can also be beneficial to the study general tendencies as some of them might not be expected. That would be the case of explicitation in brief interactive segments, similarly to what would in principle be expected in subtitles (Peguero 2003). Explicitation can be considered one of the least controversial proposed tendencies, as most studies have confirmed it (i.e. Vanderauwera, 1985; Blum-Kulka, 1986; Baker 1995, 1996; Olohan & Baker 2001; Puurtinen 2004; Whittaker 2004; Englund-Dimitrova, 2005; Kenny 2005; Saldhana 2008), while few others have found contradictory evidence (Cheong 2006; Puurtinen 2004; Grenger 2003). Web texts need to be as concise and brief as possible according to usability and style guidelines (Nielsen & Loranger 2006; Nielsen 2000; Jeney 2007). This is due to the fact that, normally, users read 25% slower on screen that on paper and therefore, usability research has lead to the recommendation to make web texts up to 50% shorter than similar printed texts. As an example, in the case of navigation menus it is recommended that they should be as brief as possible to reduce the cognitive overload in web navigation (Spyridakis 2000). Moreover, many interactive segments are allotted a specific amount of characters, normally 30% more than the source text, and any longer renderings would require the potential intervention of localization engineers (Esselink 2000). It is therefore logical to assume that explicitation, understood as "an overall tendency to spell things out rather than leave them implicit" (Baker 1996: 180), might not be generally expected in localised web texts. Another benefit of using these texts is due to the fact that most previous research has been based on literary, journalistic or popular scientific texts (Englund-Dimitrova, 2005). Some studies have also been done in subtitles (Peguero 2003), but

still, there are many types, genres or modalities that have been so far disregarded in this area of research.

Another proposed tendency is that translated texts tend to be more conventional either at the lexical, collocational or syntactic level than non-translated texts. Contrary to the case of explicitation, this tendency would in principle be beneficial to most localization processes. A cluster of proposed tendencies have focused on this, such as 'conventionalization' and 'normalization' (Baker 1993, 1997), 'sanitization' (Kenny 1998), 'simplification' (Laviosa 2002) or the 'law of standardization' (Toury 1995). Baker (1997: 183) defined conventionalization and normalization as "normalized texts display exaggerated features of the target language and conform to its typical patterns", while Kenny´s (1998) proposed sanitization was related to the use of more conventional collocations in translated texts. The definition of the law of standardization proposed by Toury (1995: 268) is also as generic as the one proposed by Baker, and it was defined as "in translation, textual relations obtaining in the original are often modified, sometimes to the point of being totally ignored, in favour of [more] habitual options offered by a target repertoire". In sum, despite different approaches all these proposed tendencies have in common that target texts have a higher tendency to conform to more conventional or typical patterns of the target language, or "what was creative and original in the source text becomes humdrum and typical in the target text" (Bowker 2001: 350). The focus is therefore clearly in the relationship between unconventional or creative features found in source texts and their reflection in target texts. The underlying model for taking up this type of research is that target texts should incorporate a similar percentage of creative or non-conventional features in the target language system. Nevertheless, in many instrumental types (Nord 1997), the opposite trend can also be found: some texts, such as websites, need to conform to the body

of conventions expected by users in order to be fully effective (Nielsen 2000; Nielsen and Tahir 2002). Therefore, not complying with the set of expected conventional features can be detrimental to the communicative and pragmatic purpose of any localised website (Nielsen 2004). Clearly, the conventionalization/normalization tendency would be in fact beneficial to the overall purpose of web localisation processes. This is supported by the fact that empirical studies have in fact confirmed that compliance to the conventions in digital web genres improve the navigation, comprehension, usability and performance (Vaughan and Dillon 2006). The research question for the study by the previous conventionalization study (Jiménez-Crespo 2009a) was therefore to test whether in fact localised texts are as conventional, if not more, that spontaneously produced websites.

The last previously reported tendency is related to the cloning of the source text structure, both at the macro and microstructural level. The methodology to study this proposed tendency is based on models borrowed from Discourse Analysis and ESP (Swales 1990; Bhatia 1993; Paltridge 1997). These models have been widely used in translation research from a contrastive perspective (Göpferich 1995; Gamero 2001), as well as digital genre theory (Crowston 2010; Askehave and Nielsen 2005; Shepherd and Watters 1998). In these models, a prototypical genre's superstructure is subdivided in recurring 'sections>moves>steps'[1] (Swales 1990) or 'communicative blocks>communicative sections>significant units>significant subunits' (Gamero 2001). Translation studies researchers have applied this type of monolingual or contrastive analysis to the study of translation products, mostly from a didactic or professional approach. In this type of research, the goal is to identify the prototypical textual structures of specific genres, while at the same time identifying the conventional linguistic forms associated with the discrete textual subunits into which any genre is subdivided. This approach can be used

to contrast the super and macrostructures of original and translated genres, thus providing evidence of structural differences between both textual populations targeted.

Three studies have dealt with this issue in localised websites (Jiménez-Crespo forthcomingb, 2010, 2009b). The first study dealt with the superstructure of localised websites, a structure that is normally represented in sitemaps and hierarchical navigation menus. Normally, localizers work with source websites with a fixed superstructure, that is, the different webpages and sections that make up the site. Contrasting the superstructures of original and localised websites can be instrumental in identifying whether or not they diverge, and why. The other two studies used different conventionalized communicative sections in corporate websites, 'contact forms' (Jiménez-Crespo 2010) and 'legal web texts', such as 'privacy policies' and 'terms and conditions' (Jiménez-Crespo forthcomingb). Both studies confirmed that localised texts showed traces source text structures, such as the inclusion of a field for *Middle Name* or *Salutation* in localized forms into Spanish when they do not appear in any of the original websites compiled. As far as the legal web texts are concerned, most privacy or e-commerce legal issues are regulated in the Spanish sociocultural context by applicable Spanish laws, while US sites are normally self regulated by companies themselves under the guidance by the Federal Trade Commission (Liu and Arnett 2000). The results of the study showed that the macro and microstructures of localised legal texts were different from those found in original Spanish ones, thus confirming the cloned text hypothesis.

Once the theoretical rational behind the reviewed studies has been explored, the next section summarized the methodology used. It will progress from the more general compilation

process of the Spanish Web Comparable Corpus, to the more specific details of the extraction of comparable subcorpora or the methodologies used for each analysis.

3. Methodology

The empirical data for most of the studies was provided by the Spanish Web Comparable Corpus, a representative collection of original Spanish and localised corporate websites from the largest US companies addressed to customers in Spain. The corpus was compiled by Jiménez-Crespo (2008a) in the context of a wider study of the description of the language of localization. Given that notions of representativeness and comparability are essential in corpus-based translation studies (Laviosa 2002; Olohan 2004; Zanettin 2000; Halverson 1998), the methodological foundation for the compilation process will be briefly described.

3.1. The Spanish Web Comparable Corpus

The corpus compilation process was divided into two parts: original and localised web texts. For the original section of the corpus, the textual population targeted was the digital genre, corporate website, in Spain.[3] The Google Business directory for Spain was used in the selection process in order to guarantee the representativeness of the resulting corpus. All directories and subdirectories were searched, and the first website from each subdirectory that was originally developed in Spanish was selected. This procedure was intended to guarantee that websites of large, medium and small companies from all business sectors in Spain were represented. The

final synchronic compilation process took place on May 4th 2006, and yielded a corpus of 172 original Spanish websites.

The localised section of the corpus includes all websites from the top 650 US companies that had a localised site for Spain on November 2006.[4]   As shown in Table 1, the selection process found 114 corporate websites specifically localised for Spain, but only ninety-five of these could be stored due to limitations of the software tool used. The resulting corpus was synchronically downloaded on the 23rd of November, 2006. Despite software limitations, the theoretical and methodological foundation of the compilation process guaranteed, to the highest possible degree, that the resulting corpus could be considered representative of both textual populations targeted.

Spanish Web Comparable Corpus	Original Section 172 websites		Localized Section 95 websites	
	Total	Average	Total	Average
Web pages	19,102	111.5 Pages/site	21,322	224.3 Pages/site
Words in body of pages	4,945,103	258.87 Words/page	8,871,512	416.07 Words/page
Words total	8,659,856	453.34 Words/page	12,562,894	589.50 Words/page

Table 1. Description of the Spanish Web Comparable Corpus

Additionally, several comparable subcorpora were extracted from the main corpus in order to test the hypothesis in the different textual elements or moves in websites. As previously mentioned, the theoretical framework behind the extraction of the different subcorpora was based on a genre approach to the study of digital texts (Jiménez-Crespo 2008b), in which a global genre is subdivided into its constituent 'moves' and 'steps' (Swales 1990). The resulting textual superstructure is represented in Figure 1. Each thematic unit in a website represented in

the navigation menu or sitemap, such as *contact us* or *about us*, is identified as a unique move in the overall structure of the hypertext (Askehave & Nielsen 2005; Jiménez-Crespo 2008c). Moreover, each move is subdivided into steps, such as the conventional *history*, *location* or *mission* pages inside the section that describes the company in corporate websites. This genre based approach is used initially to compile different subcorpora of specific communicative sections, such as contact us or legal web pages. This approach is used to study the potential tendency of translated texts to clone source text structures. For that purpose, each localised website was analyzed and all entries in navigation menus and sitemaps were assigned to a move or step in order to quantify the frequency of use. This provided a detailed statistical analysis of the frequency for all moves and steps that make up the global website superstructure. Once the textual profiles for both original and localized texts were obtained, a contrastive analysis was performed to contrast the structure of localised texts to that of spontaneously produced ones.
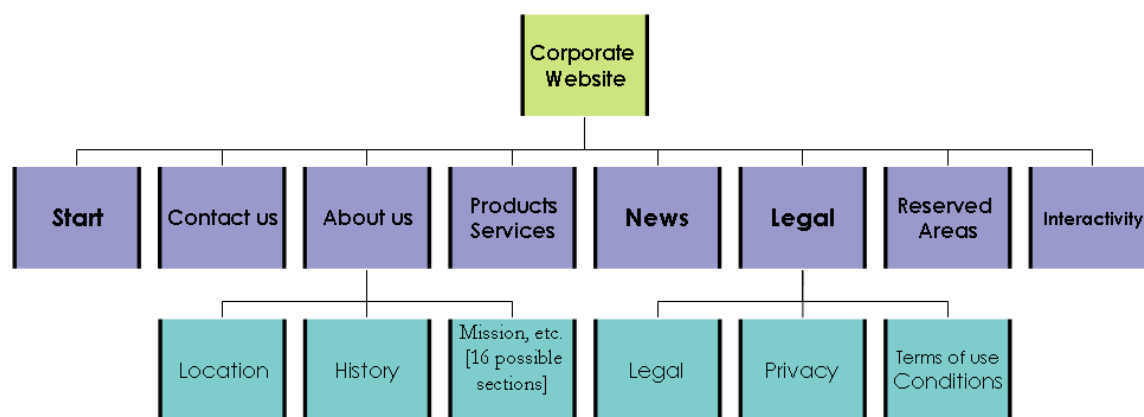

Figure 1. Corporate website prototypical superstructure and its constituent communicative blocks or 'moves' (Jiménez-Crespo 2008b; 2009a)

In addition, each individual webpage also possesses an intratextual structure in which the content is mainly divided into: (1) 'interface texts' (Price and Price 2002), brief communicative

segments whose function is to articulate the hypertextual unit, such as navigation menus, and (2) 'content texts', the unique and distinct content in each page in a website. This distinction between interface and content texts is key to the methodology used because hypertexts can potentially incorporate content texts written for other mediums. However, interface texts, such as navigation menus or sitemaps, are highly conventionalized and specific to each digital genre.

Figure 2. Distinction between 'content' and interactive 'interface' texts in websites (Price and Price 2002).

In order to analyse brief communicative segments in websites, a subcorpus that included all lexical units in navigation menus was compiled during the structural analysis stage. This corpus will be used to test the explicitation and the conventionalization hypothesis. Table 2 summarizes the characteristics of the comparable subcorpus of navigation terminology in corporate websites.

Subcorpus of Navigation terminology	Original Section		Localized Section	
	Total	Average Per site	Total	Average per site
Navigation menus	172		95	
Total lexical units	1151	6.69	1566	16.48
Words total	1762	10.24	3034	31.93

Table 2. Description of the navigation terminology corpus.

Once the main characteristics of the main corpus and the process and rationale for extraction of different subcorpora, the following section offers an overview of results.

4. Overview of results

The overview of the results will be divided in three parts, (1) a review of the studies on explicitation, including an analysis on syntactic explicitation (Olohan and Baker 2000) of both obligatory features in one language that are optional in the other, such as the case of personal pronouns that are optional in Spanish and obligatory in English, as well as syntactic explicitation of articles that are optional in both languages (Jiménez-Crespo in press); this will be followed by a summary of the results of the study by Diéguez on explicitation in terminology (2008); (2) a review of the study on conventionalization, specifically the results obtained in web navigation terminology (Jiménez-Crespo 2009a); and (3) a review of a comparative study of textual structures in original and localised websites (Jiménez-Crespo 2009b).

4.1. Web localization and explicitation (Jiménez-Crespo in press; Diéguez 2008)

The results are presented separately for each of the two stages of analysis mentioned in the previous section, followed by a summary of the results by Diéguez (2008). The optional explicitation of personal pronouns functioning as subjects was selected because they are required in English but not in Spanish. Normally, personal pronouns functioning as subject are use in Spanish for emphasis or when the subject might be unclear. The rationale for the analysis is that the syntactic explicitation hypothesis (Olohan and Baker 2000) would be supported if the frequency of use of personal pronouns in localised texts is higher than in non-translated ones. This type of explicitation could be defined as 'optional explicitation' in Klaudy´s terms (1998). For this contrastive analysis, the Wordlist function of the lexical analysis tool was used. The overall frequency list of all words in the complete overall corpus will be contrasted in both sections.

Figure 3 shows the results of the contrastive analysis using word frequency analysis. The frequency of the most repeated personal pronouns, *usted* (you-formal) (localised=0.7, original= 0.3) and *nosotros* (we) (localised=0.084, original= 0.047) is consistently higher in localised texts. The other personal pronoun that could be used in order to address the user, *tú* (you-informal), is slightly more widely used in the original tests, but the results may be biased given that in Spanish, the personal pronoun *tú* (you-informal) and the possessive *tu* (your) are typographically distinguished using an accent mark. Unfortunately, localised web texts are extremely inconsistent in their use of accentuation of monosyllabic words in Spanish (Jiménez-Crespo 2008a), and therefore, the results obtained might be biased due to localised texts not using the accent mark on this personal pronoun.
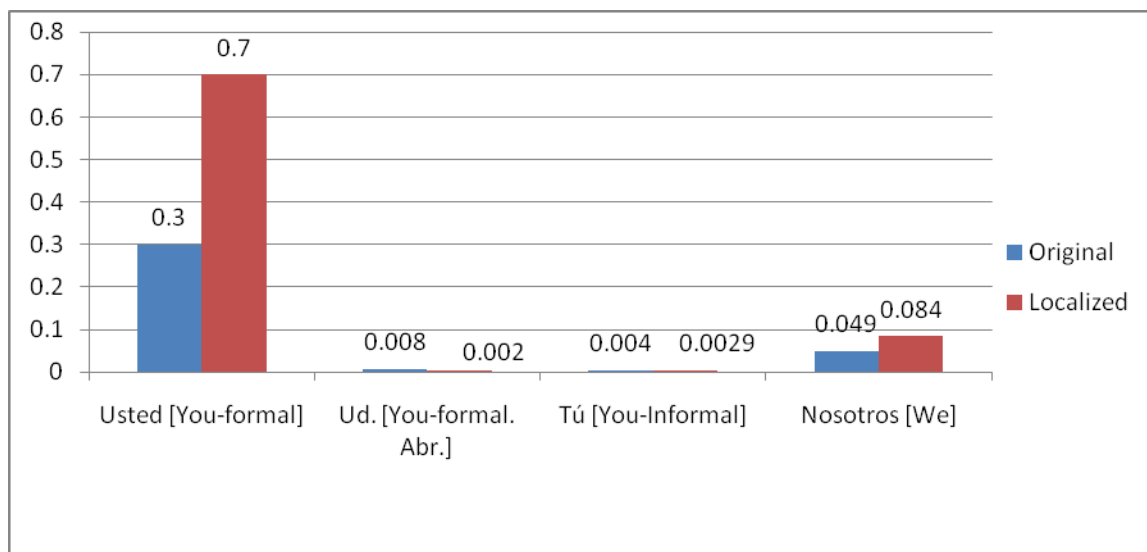
Figure 3. Contrastive analysis of personal pronoun use in the original and localised sections of the Spanish Web Comparable Corpus.

Another result of this contrastive study that supports the hypothesis is the higher frequency of use in original of the Spanish abbreviation for the pronoun *usted* (you-formal), *Ud.* This abbreviation is four times more frequent in original (0.008) than in localised (0.002), and this might indicate the tendency in originals towards language economy and brevity, a tendency not shared by localised.

For the next analysis, the comparable subcorpus of navigation terminology was used. Each entry in the navigation menus was considered as a single lexical unit as they represent concepts related to the structure of the hypertext, such as 'sitemap' or 'about us'. These structural concepts represent each of the conventionalized blocks in the digital genre 'corporate website' (Jiménez-Crespo 2009; 2008c). Several analyses were carried out, but for this review only the contrastive analysis of the optional use of the article is presented. Most lexical units in

navigation menus consist of nouns, and therefore the optional use of articles could be used to research optional syntactical explicitation.
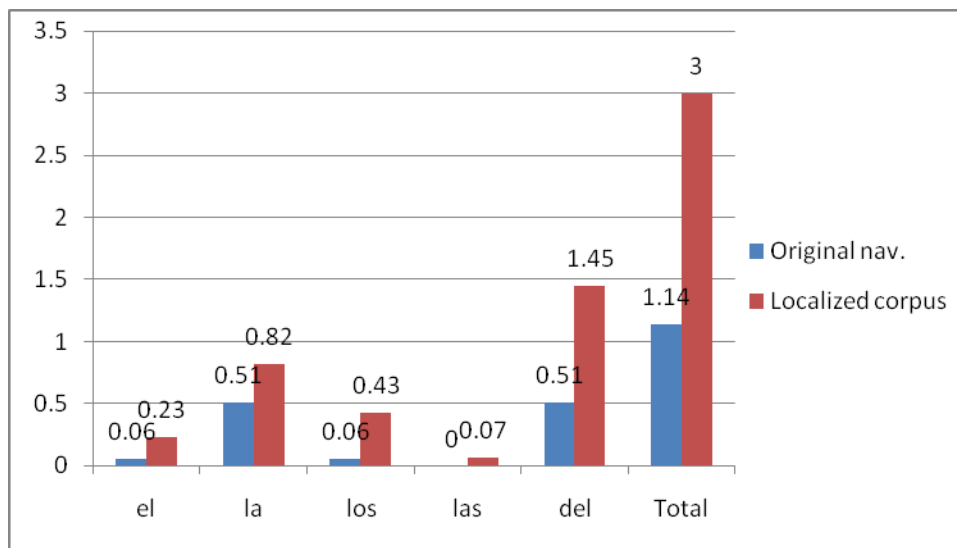

Figure 4. Contrastive analysis of article use in original and localized navigation terminology subcorpus.

Figure 4 shows this contrastive analysis for the four articles in Spanish, *el* (masculine singular), *la* (feminine singular), *los* (masculine plural), *las* (feminine plural) and overall article use. The analysis also incorporated the Spanish contraction *del* (de+el) that results from the combination of the preposition *de* (of/from) plus the masculine singular article *el* (he). This contrastive analysis clearly shows that the frequencies in localized text are consistently higher than in original ones for all articles and the contraction *del*. Globally, articles amount to 1.14% of word use in original navigation menus, while they represent 3% in localized menus. In the case of the masculine singular article, *el*, together with its prepositional contraction *del*, the frequency in localized text is almost three times higher (2.95 times) than in original texts.

The contrastive analysis summarized above demonstrate that both optional personal pronouns and articles show higher frequencies of use in localised than in original texts, even when both are in principle produced with the same textual function and addressed to the same audience. Thus, it is logical to assume that the higher tendency to explicitate the Spanish optional personal pronoun and articles in navigation menus is a more prominent tendency in localised texts.

As far as the study on explicitation by Diéguez (2008), this study analyzed 659 websites of US companies and found that 91 (13.8%) of them offered a localised version for the Chilean market. From this 91 websites translated from English into Spanish, a parallel corpus of 259 terms extracted, and the goal of the analysis was to identify whether explicitation or implicitation was preferred. Methodologically, this study is quite relevant to research on explicitation as it compares one tendency to the opposite one, implicitation. According to Toury (2004) and Klaudy and Karoly (2003), contrasting these two opposite tendencies is one of the necessary steps towards confirming any potential findings. The analysis showed that explicitation (13.2%) was more frequent than implicitation (6.1%) when rendering source English terminology into Spanish.

These studies therefore confirm a tendency towards syntactic and lexical explicitation in localized texts. The next section reviews the case of conventionalization/normalization.

4.2.   Web localization and conventionalization/normalization

The first step in order to research whether localised texts comply to the existing conventions in corporate websites originally produced in Spanish was to extract all possible lexical units in both corpora that refer to similar superstructural concepts, such as 'contact us', 'privacy policy' or

'sitemap'. As an example, Figure 5 shows a contrastive analysis of the lexical units used for the superstructural concept 'sitemap'. The analysis shows that the most frequent lexical unit in original sites is *mapa web* (54.3%), while the most frequent in localised ones is the calque *mapa del sitio* (41.7%). This last lexical unit has a frequency of only (14.3%) in original sites. The second most frequent unit in localised sites is the direct borrowing, 'sitemap' (18.3%), while this unit does not appear in original sites. It can be observed the original sites clearly favour a conventional unit, *mapa web,* while localised sites not only prefer a different unit, but also, a much greater denominative variation exists, with a 26.3% of sites using lexical units that do not appear in spontaneously produced Spanish sites.
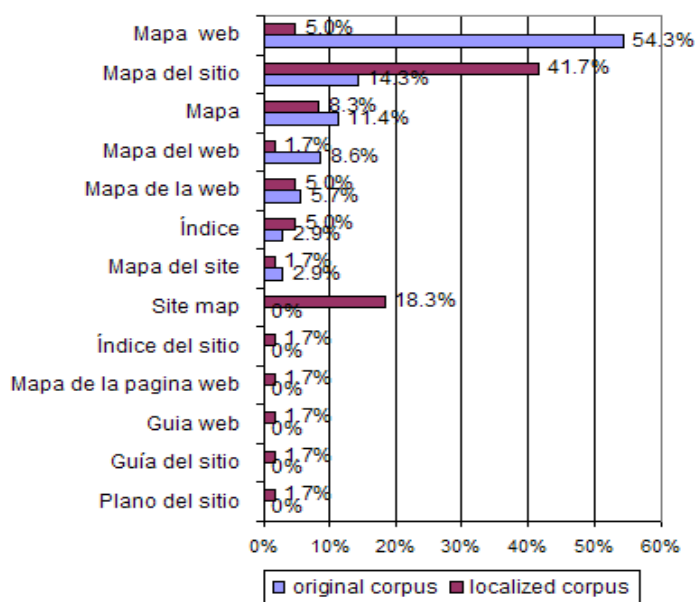

Figure 5. Contrastive analysis of lexical units used for the concept 'sitemap' in Spanish original and localized corporate websites.

A similar analysis was carried out for lexical units related to the 24 most frequent communicative blocks in original and localised corporate websites (Start, contact us, about us,

location, location, mission, quality, history of company, news, events, products, services, promotions, legal disclaimer, privacy policy, terms and conditions, clients, jobs, investors, site map, search, FAQs, links, register). In original texts, the average frequency of the most used lexical unit for the same communicative block, such as *contacto* or *empresa*, was 58.26%. Therefore, it could be deducted 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 less conventionalization 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.
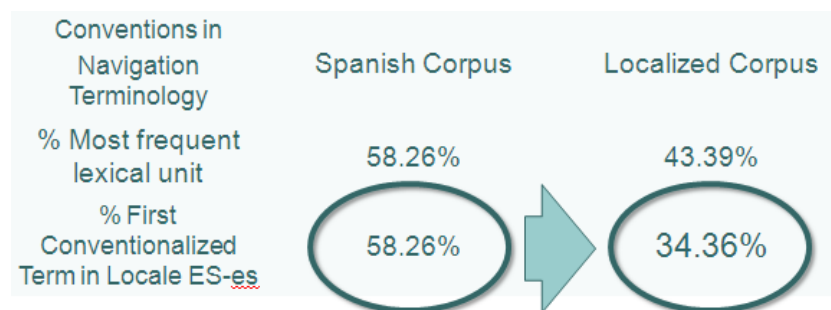

Figure 6. Summary. Average frequency of conventional lexical forms in both sections of the corpus.

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 lexical unit *mapa web* in Figure 5. In localised websites, the frequency of use of this conventional lexical unit is 5%, that is, localised websites rarely use the preferred lexical unit among those native Spanish speakers responsible for producing these sites. 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 goes down 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 the targeted discourse community.

4.3.   Web localization and cloned structures

This sections review exclusively a contrastive analysis of the superstructures of original and localised websites (Jiménez-Crespo 2009b). As previously mentioned in the 'methodology' section, a structural profile of both textual populations in the corpus, original and localised texts, were obtained and contrasted. The first significant finding of the contrastive superstructural analysis is that both textual profiles share the same number of possible moves or communicative bocks. This indicates that, to some extent, the internationalization of this web genre has led to a similar number of possible moves and steps in original Spanish sites and those localised into this language. Nevertheless, this does not mean that both superstructures are similar: the frequency of appearance for several moves, such as privacy policy or terms of use, diverges considerably.

Given that, in principle, all texts are directed towards the same target audience and sociocultural context, it is assumed that any differences between both textual profiles can be attributed directly to the replication of the source text structure.
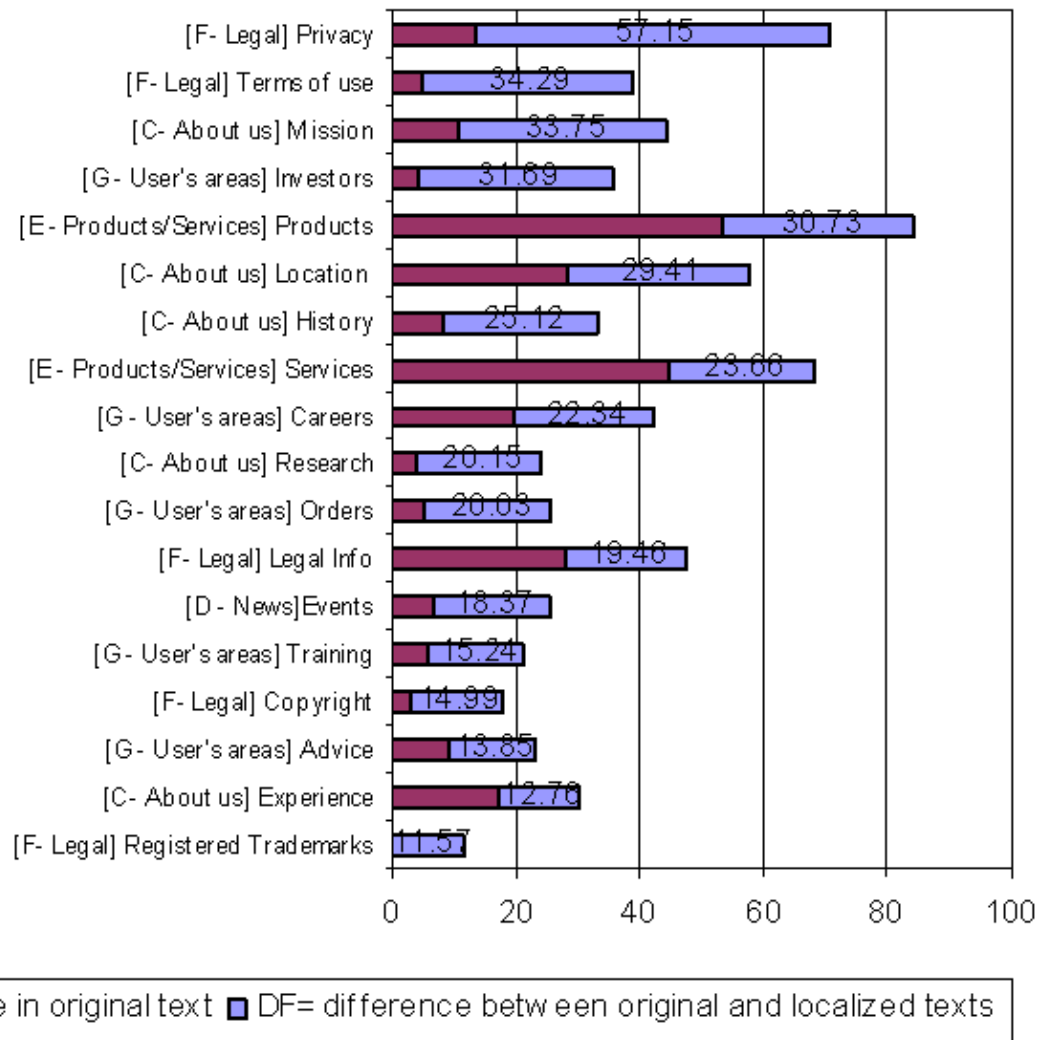

Figure 7. Superstructural differences between original and localised websites

Figure 7 shows the contrastive analysis of the frequency of appearance for each move and step. It clearly illustrates the superstructural differences between both textual profiles. It is organized according to the difference in the frequency between original and localised websites:

the red segment of each column represents the average frequency for moves or steps in original websites, the frequency of use in localised sites for the same move is represented by the total figure in each column, while the blue portion represents the variable that reflects the difference in frequency (DF) between both textual profiles.

The superstructural differences between both textual profiles are mostly concentrated in two moves: legal (F) and about us (C). In the latter move, it is of interest the higher frequency of the step that contains the values or mission of the company (DF=33.75%). This could be indicative of a conventionalized feature in US corporate websites reflecting the need to appeal to tradition and values in the US market. Nevertheless, and as shown by these results, this type of information is not conventionally offered in original Spanish sites (original=10.46%).

The most significant differences are concentrated in all moves or thematic units related to legal content, such as privacy policy (DF=57.15%), terms of use (DF=35.29%), legal information (DF=19.46%), copyright (DF=14.99%) and registered trademarks (DF=11.59%). It is fair to assume that the value of the variable DF reflects differences in the prototypical superstructure in this genre between the source and target sociocultural context, in particular, differences due to their legal systems. This finding is consistent with the results from an earlier study on corporate websites concluding that the most consistent difference between US corporate sites and other national sites was that in the former privacy webpages were more frequent (Robbin & Stylianou 2003). In fact, online privacy protection in the United States is self-regulated by companies under the guidance of the Federal Trade Commission, while this is regulated in Spain by the Spanish Data Protection Act of 1999. This means that US websites are required to explicitly formulate a full privacy policy, while Spanish sites simply indicate that their online privacy practices are in compliance with the above mentioned Spanish law. This can

explain the high frequency (localised=79%) of localised North American corporate websites including an independent privacy policy page while very few of Spanish sites do (original=10.46%). These results prove that once US sites are localized, the structure of the source text is somewhat replicated in the target text. This is consistent with what Larose (1998) refers to as 'cloned texts', that is, translated texts whose superstructure is fully maintained in the target text regardless of intercultural macrostructural differences for the same textual genre. Thus, the hypertextual page by page structure is somewhat maintained during the localization of websites, regardless of the conventional mental model of the genre structure shared by the target discourse community as represented by the websites produced by members of that community. It should be mentioned that the results of this analysis were also confirmed with the other two micro and macrostructural analyses using the same corpus (Jiménez-Crespo forthcomingb, 2010).

5. Conclusions

The purpose of this paper was to shed some light into several proposed 'general' tendencies through a summary of previous research using corpora of original and localised corporate websites. In the light of Chesterman´s (2004: 47) call to test proposed general tendencies in different translational subsets, the review of studies presented in this paper adds to the body of knowledge testing previously proposed tendencies. Additionally, a new potential tendency, 'cloned texts', has also been proposed. It represents a departure for the mainstream empirical descriptive approach brought by corpus-based translation studies as it is not based on

quantitative and qualitative lexical analyses, but rather, on a genre approach to the study of translated genres (Gamero 2001; Askehave and Nielsen 2004).

The results of previous studies summarized on this paper confirm the explicitation hypothesis in a digital genre in which, in principle, explicitation resulting in longer renderings would not be expected. Localized websites have shown traces of optional syntactic explicitation (Olohan and Baker 2000) when compared to similar spontaneously-produced texts. This tendency was observed both when optional syntactic items can be due to interference (optional personal pronouns in Spanish that are compulsory in English) and in optional syntactic items in both languages (articles in navigation menus). Moreover, the study by Diéguez (2008) also confirmed that explicitation is more frequent than implicitation in rendering source English terminology from corporate websites into Spanish.

As far as the cluster of hypothesis that suggest that translated texts favour more conventional or less creative features than non-translated texts, navigation terminology was chosen as it is highly conventionalized in most languages (Nielsen and Tahir 2002). The results from navigation menu terminology, an instrumental type of translation (Nord 1997), do not support the generality that these hypotheses entail. In fact, localised navigation terminology shows a wide range of lexical variation, and on average, the most conventional lexical unit in original sites is used on 58.26% of sites, while this same terminology is used on 34.36% of sites.

This paper also proposed conceptualizing as a general tendency a phenomenon previously observed in translated texts: the cloning of source texts structures regardless of the preferred textual structures for similar text types or genres in the target sociocultural context. The results confirmed that the structure of localised websites is somewhat different from that of

similar texts in the sociocultural context of reception. This has been observed both at the superstructural (Jiménez-Crespo 2009b) and at the macro and microstructural levels (Jiménez-Crespo forthcominga, forthcomingb).

The implications of this body of research for our understanding of the language of translation (Baker 1998; Olohan 2004) are manifold. As far as the continuous research into translational tendencies, these findings add to the body of knowledge confirming the explicitation tendency, while they do not confirm that translated texts are necessarily more conventional than non-translated ones. In fact, the translation process seems to result in more lexical variation, therefore producing many non-conventional lexical units when the target language has a preferred convention in a given genre to express a recurrent communicative form. Additionally, in line with the many limitations inherent to corpus-based translation research (Baker 2004; Tymoczko 2005), the tendency to copy source text structures resulting structural differences between translated and non-translated texts shows that the scope of translational tendencies should not be limited to analyses using lexical tools.

As far as the practice of translation, the objectives of the industry for localization process is that localized websites should "look like they have been developed in-country" (LISA 2004:11). Therefore, professionals localisers and quality testers can be aware of these tendencies and "consciously chose [or not] to resist them" (Chesterman 2004: 46). Moreover, future localizers that are currently being trained could clearly benefit from an understanding of translation tendencies. Trainers can direct the attention of students towards these tendencies, instead of replicating currently established processes that contribute to less conventional or

cloned structures, as well as include parallel and comparable corpus research as a tool during the translation process.

In the continuing evolution of Translation Studies, the impact of technology as a revolutionary force cannot be underestimated. As Munday (2008:179) indicated: "The emergence of new technologies has transformed translation practice and is now exerting an impact on research and, as a consequence, on the theorization of translation". It is expected that this review will both assist in the continuing search for generalities in translational behaviour, as well as potentially encourage new research into the impacts and effects of technology on the discipline.

Notes

[1] For our purposes, a *move* is defined as a "unit of discourse structure which presents a uniform orientation, has specific structural characteristics and has a clearly defined function" (Swales 1990: 140).

References

Alcina, A. (2008). "Translation technologies: scope, tools and resources". *Target,* 20, 79-102.

Askehave, I. and A. E. Nielsen (2005). "Digital genres: a challenge to traditional genre theory".
    Information Technology and People, 18, 120-141.

Austermühl, F. (2001). *Electronic tools for translators*. Manchester: St Jerome Publishing.

Baker, M. (1993). "Corpus Linguistics and Translation Studies: Implications and Applications". In M. Baker, G. Francis and E. Tognini-Bonelli (eds.) *Text and Technology: In Honour of John Sinclair*. Amsterdam-Philadelphia: John Benjamins, 233-250.

Baker, M. (1995). "Corpora in Translation Studies: An overview and some suggestions for future research". *Target,* 7, 223-243.

Baker, M. (1996). "Corpus-based Translation Studies: The Challenges that Lie Ahead". In H. Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam-Philadelphia: John Benjamins, 175-186.

Baker, M. (1997). "Corpus-based Translation Studies: The Challenges that Lie Ahead". In H. Somers (ed.) *Terminology, LSP and Translation: Studies in Language Engineering in Honour of Juan C. Sager*. Amsterdam, John Benjamins, 175-186.

Baker, M. (1999). "The role of corpora in investigating the linguistic behaviour of professional translators". *International Journal of Corpus Linguistics,* 4, 281–298.

Baker, M. (2004). "A Corpus-based View of Similarity and Difference in Translation". *International Journal of Corpus Linguistics,* 9, 167-193.

Bowker, L. (2006). "Translation memory and 'text' ". In L. Bowker (ed.) *Lexicography, terminology and translation*. Ottawa: University of Ottawa Press, 175-187.

Bowker, L. (2001). "A Corpus-Based Approach to Translation Evaluation". *Meta, 46,* 345-363.

Bowker, L. (2002). *Computer-Aided Translation Technology: A Practical Introduction*. Ottawa: University of Ottawa Press.

Blum-Kulka, S. (1986). "Shifts of cohesion and coherence in translation". In J. House and S. Blum-Kulka (eds.) *Interlingual and Intercultural Communication: Discourse and*

Cognition in Translation and Second Language Acquisition Studies. Tübingen: Gunter Narr, 17-35.

Cheong, H.J. (2006). "Target text contraction in English into Korean Translation: a contradiction of presumed translation universals?". *Meta,* 51, 343-367.

Chesterman, A. (2010). "Response and Discussion". Paper presented at the *MATS 2010, Methodological Advances in Corpus-Based Translation Studies Conference,* University College Ghent, Belgium, January 8th, 9th, 2010.

Chesterman, A. (2004a). "Beyond the Particular". In A. Mauranen and P. Kujamäki (eds.) *Translation Universals: Do they exist?*. Amsterdam-Philadelphia: John Benjamins, 33-50.

Chesterman, A. (2000). "Causal Model for Translation Studies". In M. Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects.* Manchester: St. Jerome Publishing, 15-27.

Cronin, M. (2003). *Translation and Globalization*. London: Routledge.

Crowston, K. (2010). "Internet Genres". *Encyclopedia of Library and Information Sciences, Third Edition, Vol IV*. New York/London: Routledge, 2983-2995.

Dieguez Morales, M. I. (2008). "Análisis terminológico de sitios web localizados del inglés al español: uso de técnicas de amplificación y elisión", *Tradumática* 6. Available at: http://www.fti.uab.es/tradumatica/revista/num6/articles/09/9art.htm. (accessed: 27 September 2010).

Englund-Dimitrova, B. (2005). *Expertise and Explicitation in the Translation*. Amsterdam-Philadelphia: John Benjamins.

Esselink, B. (2000). *A Practical Guide to Localization*. Amsterdam - Philadelphia: John
Benjamins.

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

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

Halverson, S. L. (1998). "Translation studies and representative corpora: establishing links between translation corpora, theoretical/descriptive cathegories and a conception of the object of study". *Meta,* 43, 494-514.

House, J. (2008). "Beyond intervention: universals in translation". *Trans-kom*, 1, 6-19.

Jiménez-Crespo, M. A. (2008a). *El proceso de localización web: estudio contrastivo de un corpus comparable del género sitio web corporativo*. Department of Translation and Interpreting, University of Granada, Spain. [PhD thesis]. Available at: http://hera.ugr.es/tesisugr/17515324.pdf (Accessed: 27 September 2010)

Jiménez-Crespo, M. A. (2008b). "Web genres in Localization: a Spanish corpus study". *Localization Focus – The International Journal of Localization,* 6, 2-13.

Jiménez-Crespo, M. A. (2008c). "Caracterización del género «sitio web corporativo» español: análisis descriptivo con fines traductológicos". In M. Fernández Sánchez and R. Muñoz Martín (eds.) *Aproximaciones cognitivas al estudio de la traducción y la interpretación*. Granada: Comares, 259-300.

Jiménez-Crespo, M. A. (2009a). "Conventions in localisation: a corpus study of original vs. translated web texts". *Jostrans: The Journal of Specialized Translation*, 12, 79-102.

Jiménez-Crespo, M. A. (2009b). "The effect of Translation Memory tools in translated web texts: evidence from a comparative product-based study". *Linguistica Antverpiensia* 8, 213-232.

Jiménez-Crespo, M. A. (2010). "The intersection of localization and translation: A corpus study of Spanish original and localized web forms". *Translation and Interpreting Studies,* 5, 186–207.

Jiménez-Crespo, M. A. (Forthcominga). "What the future holds for the study of general tendencies in translation: explicitation in web localization". *Target*, 23.

Jiménez-Crespo, M. A. (Forthcomingb). "To adapt or not to adapt in web localization: a contrastive genre-based study of original and localized legal sections in corporate websites" *Jostrans: The Journal of Specialized Translation*, 14.

Klaudy, K. (1998). "Explicitation". In M. Baker (ed.) *Routledge Encyclopaedia of Translation Studies*. London: Routledge, 80–84.

Klaudy, K. and K. Karoly. (2005). "Implicitation in translation: An empirical justification of operational asymmetry in translation". *Across Languages and Cultures,* 6, 13-28.

Larose, R.. (1998). "Méthodologie de l'évaluation des traductions". *Meta* 43:2. 163-186.

Laviosa, S. (1998). "The Corpus-based Approach: A New Paradigm in Translation Studies". *Meta*, 43, 474–479.

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

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

Liu, Chang and K. P. Arnett (2000). "An examination of Privacy Policies in Fortune 500 Web Sites". *Mid-American Journal of Business,* 17, 13-21.

Mauranen, A. (2007). "Universal tendencies in translation". In M. Rogers and G. Anderman (ed.) *Incorporating corpora: the linguist and the translator.* Clevedon: Multilingual Matters, 33-48.

Munday, J. (2009). *The Routledge Companion to Translation Studies*. London-New York: Routledge.

Nielsen, J. (2000). *Designing Web Usability: the practice of simplicity*. Indianapolis: News Riders.

Nielsen, J. (2004). "The Need for Web Design Standards", *Jakob Nielsen's Alertbox,* September 13, 2004, http://www.useit.com/alertbox/20040913.html.

Nielsen, J. and H. Loranger (2006). *Prioritizing Web Usability*. Indianapolis: News Riders.

Nord, C. (1997). *Translating as a Purposeful Activity. Functionalist Approaches Explained.* Manchester: St. Jerome.

Nielsen J. M. and M. Tahir. (2002). *Homepage usability: 50 Websited deconstructed.* Indianapolis: News Riders.

Olohan, M. and M. Baker. (2000). "Reporting that in translated English: Evidence for subconscious processes of explicitation?". *Across Languages and Cultures,* 1, 141-158.

O'Hagan, M. (2009). "Evolution of User-generated Translation: Fansubs, Translation Hacking and Crowdsourcing". *The Journal of Internationalization and Localization,* 1 94-121.

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

Perego, E. (2003). "Evidence of explicitation in subtitling: Towards a categorisation". *Across Languages and Cultures,* 4, 63–88.

Price, J. and L. Price. (2002). *Hot text. Web Writing that works.* Berkeley: News Riders.

Puurtinen, T. (2004). "Explicitation of clausal relations A corpus-based analysis of clause connectives in translated and non-translated Finnish children's literature". In A. Mauranen and P. Kujamäki (eds.) *Translation Universals: Do they exist?*. Amsterdam-Philadelphia: John Benjamins, 165-176.

Pym, A. (2010). *Exploring Translation Theories*. New York/London: Routledge.

Reineke, D. (2004). *Traducción y Localización*. Las Palmas de Gran Canaria: Anroart Ediciones.

Robbins, S. and A. Stylianou (2003). "Global corporate web sites: an empirical investigation of content and design". *Information & Management* 40, 205-212.

Saldhana, G. (2008). "Explicitation revisited: Bringing the reader into the picture". *Trans-kom,* 1, 20-35.

Schäler, R (2005). "The Irish Model in Localization". Paper presented at LISA Forum Cairo 2005: Perspectives from the Middle East and Africa. Available at http://www.lisa.org/utils/getfile.html?id=61136686. (Accessed: 27 September 2010)

Snell-Hornby, M. (2006). *The Turns of Translation Studies*. Amsterdam-Philadelphia: John Benjamins.

Spyridakis, J. H. (2000). "Guidelines for Authoring Comprehensible Web Pages and Evaluating their Success". *Technical Communication,* 47, 359-382.

Swales, J. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Toury, G. (2004). "Probabilistic explanations in translation studies: welcome as they are, would they qualify as universals?". In A. Mauranen. and P. Kujamäki (eds.) *Translation Universals: Do They Exist?* Amsterdam-Philadelphia: John Benjamins, 15-32.

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

Tymoczko, M. (2005). "Trajectories of Research in Translation Studies". *Meta,* 50, 1082-1097.

Vanderauwera, R. (1985). *Dutch Novels Translated into English: the transformation of a 'minority' literature*. Amsterdam: Rodopi.

Vaughan, M. W. and A. 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

Venuti, L. (2005). "Translation, History, Narrative". *Meta,* 50, 800-816.

Whittaker, S. (2004). "Étude contrastive des syntagmes nominaux démonstratifs dans des textes traduits du français en norvégien et des textes sources norvégiens: stratégie de traduction ou translationese?". *Forum,* 2, 221-240.

Zanettin, Federico (2000). "Parallel Corpora in Translation Studies: Issues in Corpus Design and Analysis". In M. Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I: Textual and Cognitive Aspects*. Manchester: St. Jerome, 105-118.