The evaluation of pragmatic and functionalist aspects in localization: towards a holistic approach to Quality Assurance

Miguel A. Jiménez-Crespo

Department of Spanish and Portuguese
Rutgers, The State University of New Jersey

Abstract

Current localization QA models are based on componential error-based approaches to quality evaluation. However, as it is more complex to objectively measure pragmatic issues than language or functionality problems, the communicative-pragmatic adequacy of the target text is normally ignored while assuring that a text "look[s] like it has been developed in-country" (LISA, 2004, p. 11). Consequently, most QA models do not incorporate a category for this type of issue. This paper presents a critical review of the notion of quality in localization and its implications in current QA practices from a functionalist perspective (Nord, 1997). The main goal is to set the foundation for an evaluation process that can account for functionalist and pragmatic inadequacies through the use of localization evaluation corpora. The main issues and applications are illustrated through examples extracted from the 40,000 webpage Spanish Web Evaluation Corpus compiled by the author (Jiménez-Crespo, 2008a).

DOI: 10.1075/jial.1.03jim
For over two decades, the production of digital content such as websites, software products, or videogames has increased exponentially. This digital revolution, in the context of a wide globalization process, has led to a dramatic increase in the amount of digital texts that are translated around the world (Cronin, 2003). The localization industry emerged as a response to this challenge, and it developed in a rapidly changing environment that demanded the translation of massive amounts of digital texts rapidly, efficiently, and cost-effectively. However, as some scholars have pointed out, this fast pace has meant that the industry established its practices without fully relying in the body of knowledge of related disciplines, such as Translation Studies, Interlinguistic communication, Discourse Analysis, Contrastive Rhetoric, etc. (Dunne, 2006a; Pym, 2003).

This relative lack of dialogue between the Localization Industry and Translation Studies has been problematic as certain established practices, such as current localization evaluation, are not based on explicitly formulated theoretical models or solid empirical research. In fact, the industry’s approach to evaluation has been characterized as "experienced-based or anecdotal" (Colina, 2008), a procedural approach that, given that it is based on evaluators’ previous experiences, can be subjective and, therefore, might not provide a sufficient degree of validity and reliability.

Moreover, current QA models do not implicitly incorporate the pragmatic and functional adequacy of the target text, even when this category appears in most translation error typologies.

---

1 The localization industry traditionally consisted of software and hardware manufacturers and the companies that produced localized versions, the so-called localization service providers. Nowadays, since the creation of the WWW, this industry also consists of telecommunication companies, web service providers and web development companies (Dunne, 2006a), as well as localization training institutions such as the Localization Research Centre at Limmerick, Ireland.

2 It could be argued that some other economic sectors, such as the healthcare industry in the US, have also ignored to some extent the body of knowledge of translation research.
Pragmatic and Functionalist Aspects in Localization

(Melis & Hurtado, 2001). This means that QA systems might not evaluate whether the pragmatic configuration of the target text is similar to those that “have been developed in-country” (LISA, 2004). In this context, the objective of this paper is to contribute to the foundation for a QA model that can account for these recurrent inadequacies found in published localized texts. The empirical data is provided by the contrastive analysis of a 40,000 web page comparable corpus (Baker, 1995) of original Spanish corporate websites and localized sites from the largest US companies with sites addressed to customers in Spain. In this analysis, the notions of corpus-based evaluation (Bowker, 2001) and holistic translation assessment (Colina, 2008; Waddington, 2001) play an essential role. It will be argued that a combination of these notions could lead towards a more objective base that measurement and assessment scholars (Harper, M., O'Connor, K. & Simpson, M., 1999), as well as Translation Studies scholars (Colina, 2008; House, 1997, 2001; Melis & Hurtado, 2001;) have deemed essential for evaluation processes to be valid and reliable. It should be noted that the scope of this paper is not to propose a localization evaluation model or adapt pragmatic evaluation proposals that are generally too complex for professional settings (i.e. House, 1997), but rather to direct attention to recurring pragmatic issues in real localized texts that are difficult to detect applying current quality metrics. In doing so, a more comprehensive approach to evaluation in localization will emerge. This approach will be based on explicitly formulated theoretical models and deduced from a principled collection of texts subject to current quality evaluation practices.

3 In Translation Studies, a comparable corpus can be defined as “a structured electronic collection of texts originally written in a particular language, alongside texts translated into that same language” (Baker, 1995: 234).
Quality Evaluation in Localization

Recently, the evaluation of quality in localization has attracted the attention of a rising number of scholars from different perspectives, such as Translation Studies (Bass, 2006; Dunne, 2006b; García, 2008; Jiménez-Crespo, 2008a; Pierini, 2007;), cognitive science (Börjel, 2007) and web content management (Gibbs & Matthaiakis, 2006). These studies indicate the need for further research into this phenomenon, in part because the same set of evaluation criteria cannot be uniformly applied to all translation activity (e.g. Honing, 1998, p. 14; Larose, 1998, p. 164; Melis & Hurtado, 2001, p. 284). Nevertheless, the fact that localization is relatively young, along with the relative lack of descriptive empirical research into this phenomenon, imply that it is too early to identify a conventional set of canonized criteria for evaluating localization as a distinctive translation modality (Wright, 2006). As mentioned previously, several scholars have indicated that the industry's efforts have been carried out without fully relying on the body of knowledge of Translation Studies (Dunne, 2006a; Pym, 2003; O'Hagan & Ashworth, 2003). The lack of clear theoretical basis is somewhat problematic because, as Julianne House (1997) rightly pointed out:

Evaluating the quality of a translation presupposes a theory of translation. Thus different views of translation lead to different concepts of translational quality, and hence different ways of assessing it. (p. 7)

This implies that the notion of evaluation in localization might be somewhat different from the same notion in Translation Studies, despite that fact that an agreed upon model does not even exist in the latter.

If industry QA practices are seen in the light of translation evaluation, their goal is usually summative in nature, or in other words, they are used to determine the end result or pass a final judgment (Melis & Hurtado, 2001). The other two current theoretical approaches to
translation evaluation are the diagnostic and the formative. The main differences between these
models can be found in the goals of the evaluation process: formative evaluation usually requires
providing constructive feedback from the trainer to the evaluatee (Kussmaul, 1995), while
diagnostic evaluation is directed towards level placing. The summative approach used by the
industry is mostly directed towards passing a judgment on a pass/fail basis through an error-
based assessment method. The LISA QA model is an example of this last approach, and given
its dominant position in this sector, it will be reviewed later in the paper.

Any analysis of pragmatic issues has to be contextualized in the wider context of the
localization industry and Translation Studies. It is therefore necessary to review current
implemented practices in the industry and the underlying notions related to them, such as
"quality", "errors" or the "goals" of the evaluation process.

The Notion of "Quality" in the Localization Industry

The quest towards achieving high levels of quality in all entrepreneurial processes has
found a reflection in this industry, with "quality receiving more attention than ever" (Bass, 2006,
p.6). Currently, it is assumed that a customizable componential QA model with an error-based
acceptability threshold can guarantee a sufficient degree of quality, regardless of the genre, text
type, register or display device.

If the industry's literature on this issue is analyzed, it can be deduced that most
international standards define quality as the capacity to comply with a set of parameters pre-
defined by the customer. For example, the ISO 9000 defines quality as: "the totality of features
and characteristics of a product or service that bears on its ability to satisfy stated or implied
needs" (ISO 9000, apud Ørsted, 2001, p. 443). With a very similar perspective, TQM (Total
Quality Management) defines quality as "fully satisfying agreed customer requirements". In the same line, the definition laid out by the ASTM standard defines quality as "the degree of conformance to an agreed upon set of specifications” (ASTM, apud Melby, 2006). Nevertheless, it would be theoretically and methodologically impossible to predefine the notion of "quality" in all translated texts: for this reason, common definitions of quality usually focus on procedural aspects as opposed to establishing what could be considered a "quality" translated text. Basically, such definitions govern procedures for achieving quality, rather than providing normative statements about what constitutes quality (Melis & Hurtado, 2001, p. 274). They are generically process-oriented instead of product-oriented (Corpas, 2006; Wright, 2006, p. 256). As a result, the final decision about quality resides in a time-constrained evaluation process carried out by one or more evaluators that might lack the necessary theoretical framework in order to separate their own subjective judgment (House, 2001). Following previous empirical studies of translation quality evaluation (Colina, 2008; Nobs, 2006), this paper asserts that functionalist approaches could provide a valid framework to contextualize QA models.

**Quality and a Functionalist View of Localization**

The operative quality definitions used in the industry are consistent with functionalist approaches to translation (Holz-Mänttäri, 1984; Nord, 1997; Reiss & Vermeer, 1984). In this theoretical framework, the main determining factor of the translation process is the purpose or *Skopos* of the translational action (Nord, 1997, p. 27). The source text is no longer the determinant factor in the translation or evaluation process, as the function of the target text and the context of reception are the yardsticks to which all translation decisions should be evaluated. Thus, a target text is assumed to be "functional" when it relates to the "expectations, needs,
previous knowledge and situational conditions" (ibid, p. 28) of the receiver for whom it is intended. In order to determine the functional configuration of the target text, functionalist theorists introduced the notion of "translation brief", a summary of basic aspects such as intended function(s), situational aspects, intended audience, medium, place, etc. (Nord, 1997).

These aspects are assumed to help guide the translator's decisions as well as the evaluator's judgments.

If the above mentioned quality definitions are examined in more detail, they are specifically consistent with the revision of the earlier skopos theory by Christiane Nord (1997), who presented a "function-plus-loyalty" model. In this proposal, the translator as an intercultural communicator should not merely focus on the function of the target text in the sociocultural environment of reception, but be loyal to both the skopos of the translation and the requests of the initiator, as stated in the translation brief. Nevertheless, if this approach is fully implemented, the emphasis that quality definitions place on the demands of the initiator or client has to be balanced with situational and functional aspects of the context of reception.

Additionally, as the most repeated criticism to functionalist approaches states, initiators might lack the knowledge and understanding of the translation phenomenon, rendering them unable to set a requested predefined notion of quality. In localization, this paper agrees with Dunne (2006b) in that:

> Clients 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). (p. 100)

---

4 From a functionalist perspective, an initiator can be defined as "the person or group of people or institution that starts off the translation process and determines its course by defining the purpose for which the target text is intended" (Nord, 1997: 139).
In this sense, it is problematic to establish and measure a required level of "quality" in an objective and valid manner if different participants do not share the same notion. In localization, it has been suggested that localization researchers should direct their efforts towards educating initiators on how to achieve and measure an objective level of "quality" (Bass, 2006), but again, this could only be done once empirical research into this phenomenon can provide a clearer objective foundation. In order to establish the objective of the evaluation process and the function of the localized texts, a review of the industry literature can be revealing.

The translation "Skopos" and the Objectives of the Evaluation Process

At this point, it is logical to assume that "quality evaluation" entails a balance between a notion of quality that is soundly based on translation theory (House, 2001), empirically tested methods (Colina, 2008; Rothe-Neves, 2005), and the "objectives" set forth by the initiators that request the translation (the translation brief). The validity of the process, or in other words, "the extent to which an assessment measures what it is designed to measure" (Harper et al., 1999, p. 49), thus depends on a clear definition of what is actually being evaluated. From industry publications, it can be deduced that the objective of the localization process should be producing quality texts "with the look and feel of locally made products" (LISA, 2003, p. 5), or products "that look like [they have] been developed in-country" (LISA, 2004, p. 11). In its turn, the objective of the evaluation process would be to guarantee that these texts do look like local productions and serve their purpose efficiently. Therefore, it is fair to argue that, apart from eradicating any transfer and language errors in the target language, a functional target text should "look" like texts originally produced in the target region. In order to accomplish this goal, compliance with existing conventions in natively produced texts in the target culture becomes an
essential aspect of quality (Jiménez-Crespo, 2009; Nielsen & Tahir, 2002). Conventions are shared by a discourse community and frame the expectations from receivers and possible effects of the texts on them. Conventions are culture-dependent and may differ from culture to culture (Nord, 1997, p. 54); this is the reason why they play an important role during translation processes. In localization, target texts are presented as original productions, that is, receivers do not need to be implicitly aware that they are interacting with a translated text. Their main goal is to produce in the reader an effect as close to that of a reader of the original text (Newmark, 1981, p. 39), a type of translation that has been referred to as instrumental (Nord, 1997), covert (House, 1997) or communicative translation (Newark, 1981). Thus, translators in principle would replace any conventional feature in the original text with the conventional feature in the same textual genre in the target text. For the purposes of QA analysis, this means that interlinguistic studies centered on describing conventional features in different digital genres and locales could be highly beneficial (Jiménez-Crespo, 2009, 2008a; Adkisson, 2002; Nielsen & Tahir, 2002). As an alternative approach, companies such as Facebook, Microsoft or Symantec have already implemented novel ways to identify conventional features shared by target discourse communities. This is the case of "crowdsourcing", or in other words, letting communities of web users decide and vote on the best translations for terms and phrases (O’Hagan, 2009).

Once the notion of quality has been reviewed from a functionalist perspective, the next step would be to review the ill-defined notion of "error". Functionalist researchers normally downplay the importance of errors in any evaluation process, as they assert that error analysis per-se is insufficient to evaluate the quality of a translation. Instead, "it is the text as a whole

---

5 The notion of convention can be defined as: "Implicit or tacit non-binding regulation of behavior, 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).
whose function(s) and effect(s) must be regarded as the crucial criteria [...]" (Nord, 1991, p. 166). Nevertheless, it is obvious that errors play an essential role in QA models and will continue to do so. A review of this concept is therefore needed.

**The Role of Errors in Quality Evaluation**

Errors and error typologies are the foundation upon which the QA evaluation processes are built. Normally, they include pre-established error typologies, but again, a clear and operative notion of error is not provided. In Translation Studies, the notion of error has been defined from a cognitive and functionalist perspective as a translation problem that has not been solved or that has not been solved appropriately (Nord, 1996, p. 96-100). Translation errors are therefore related to translation problems, and they can be defined as "an objective problem which every translator [...] has to solve during a particular translation task" (Nord, 1991, p. 151). After a thorough review of the notion of error in previous literature, Hurtado Albir (2001, p. 305-306) categorized them according to three possible principles: (1) errors caused by a misunderstanding of the source text, such as false sense, omission, no sense, etc., (2) language errors in the target language, such as punctuation, grammar, lexical or style errors, (3) and pragmatic errors, such as those related to inadequacies as far as the function or *skopos* of the translation is concerned (Reiss & Vermeer, 1984). These pragmatic errors can only be identified by "comparing the source and target text in the light of the translation brief" (Nord, 1997, p. 76), and native speakers reviewing exclusively the target text might not be able to identify them. This is due to the fact that pragmatically erroneous expressions might not be inadequate in themselves; they only become inadequate "with regard to the communicative function it was intended to serve".
supposed to achieve" (ibid, p. 73). These are normally more serious errors, as receivers do not realize that they are getting wrong or inadequate information. This contrasts with general language or cultural errors that can be more easily identified and corrected (Colina, 2008). Current localization error typologies do not explicitly incorporate a category for pragmatic inadequacies, and as such, QA and QC processes cannot trace the actual cause of the error and try to put a mechanism in place to correct it (Bass, 2006). It is our contention that localization QA models are in dire need of adding this third category. This could be implemented through a combination of corpus-based approaches to evaluation, the training of evaluators and/or the adoption of a holistic system (Colina, 2008; Waddington, 2001).

In order to illustrate these types of errors, the following list presents recurrent pragmatic inadequacies in published localized web texts into Spanish. It was extracted from the multilevel corpus analysis carried out by the author in his doctoral dissertation (Jiménez-Crespo, 2008a, p. 536-566). These errors were found in a corpus of 95 published localized websites from the largest US companies according to the Forbes list. These sites were contrasted with 172 original Spanish corporate sites following a corpus-based methodology:

1. More direct formulation of the appellative function in Spanish localized texts. As an example, the concordance analysis of the verb "leer" ‘read’ shown in Figure 1 illustrates how the infinitive form of the verb is more frequent in Spanish original privacy pages to address the user, while imperative form of this verb, such as "lea/lee" is overwhelmingly used in localized texts (81.81%).
Figure 1. Extract from a KWIC (Key Words in Context) display of the results of a query of the verb "leer" in original and localized Spanish web privacy policies.

This illustrates that pragmatically, Spanish texts tend to use verbal periphrasis to formulate this recurring unit of meaning. Consequently, the frequent use of direct imperative forms of verbs in texts localized into Spanish could be considered a pragmatic inadequacy, such as "¡MUY IMPORTANTE! LEA ESTE ACUERDO LEGAL [sic] 'Very important! Read this legal agreement!'". As the concordance lines shown in Figure 1 demonstrate, the user is rarely addressed directly in the imperative form of the verb in original Spanish texts, but rather with verbal periphrasis "le rogamos que lea el siguiente acuerdo" 'we encourage you to read the following agreement' (Jiménez-Crespo, 2008a). In these cases, even when the translated utterance would not be erroneous per-se, the target text would not comply from a pragmatic perspective with the expectations of members of the target discourse community.

2. Not complying with sociocultural norms: 33% of localized contact forms in the corpus under study use the singular form of the word ‘last name’, "apellido", even when Spanish users always identify themselves with two last names, "apellidos" (Jiménez-Crespo, 2008a).
3. Explicitation in brief textual segments that need to be concise due to space constraints (Jeney, 2007). The translation of "contact us" for longer phraseological units such as "Póngase en contacto con nosotros" - "Entre en contacto con nosotros" ‘get in touch with us’, although correct in Spanish, would be inadequate as they do not comply with the specific constraints for this text type.

4. Using non-conventional linguistic forms associated with the translated textual genre. A specific textual genre and each of its thematical units might incorporate recurring conventional linguistic forms (Glässer, 1990; Göpferich, 1995; Gamero, 2001), such as "contact us" or "about us" in web navigation menus (Nielsen & Tahir, 2002). A descriptive study of conventional terminology in navigation menus was carried out in both components of the evaluation corpus (Jiménez-Crespo, 2009). It provided a percentage of use for all lexical units found in original navigation menus. As an example, the term "contacto" ‘contact’ was found to be the most conventional term in original navigation menus (49.9%), while in English texts "contact us" appears in 89% of websites (Nielsen & Tahir, 2002). In original websites, the average frequency of the most used lexical unit represented in navigation menus was 58.26%. The average frequency for the most used lexical units in original Spanish texts in navigation menus, such as "contacto" or "aviso legal", is 34.36% in localized texts (Jiménez-Crespo, 2009).⁷

After an initial research stage, these types of pragmatic errors could be incorporated to QA systems. But how can they be incorporated in current established systems such as LISA's? A review of the role of errors in the LISA QA model can shed some light into this issue.

---

⁷ See Jiménez-Crespo (2009) for a complete analysis.
Errors and Pragmatic Issues in the LISA QA Model

The LISA QA analysis is the most widely used evaluation model in the localization industry (Lommel, 2008). In this error-based system, the evaluator(s) set(s) a threshold for the number of combined minor, major and critical errors that are allowed, and if that pre-defined threshold is passed, the localized text would be deemed "unpublishable". Nevertheless, what counts as a minor, major or critical error is rarely described, an aspect that can significantly increase the variability among raters. This approach also presumes that a limited number of errors is unavoidable due to the economic context in which this process is performed (Bass, 2006). In addition, even when this type of evaluation can be defined as summative in nature, it also incorporates a formative component given that evaluators might be requested to provide feedback to translators in order to minimize future errors (Esselink, 2000), or they might be requested to propose solutions to all identified inadequacies (Chandler, 2004).

As with any localization evaluation, the LISA QA model incorporates both a linguistic and a functionality testing stage (Esselink, 2000; Chandler, 2004). Both phases are usually preformed by different evaluators or testers. Nevertheless, these stages cannot be understood in isolation, and even when the linguistic testing might measure an outstanding level of quality, the overall localized text would not achieve this degree of conformance to a quality standard if it contains functionality problems, such as deficient text integration or textual segments in other

---

8 Nevertheless, it is assumed that a process should be in place to minimize and effectively correct those errors.

9 Linguistic testing deals with all the textual aspects of the evaluation process, and it could be compared to traditional translation testing. It can be defined as a stage that “focuses on all language-related aspects of a localized application” (Esselink, 2000: 150). Functionality testing deals with all the programming and functional aspects of a software program or website, such as the presence of programming errors, the inclusion of other languages in a localized version or that specific character encodings are correctly displayed.

10 In software testing Esselink (2000: 145-155) identifies internationalization testing, localization testing, functionality testing and delivery testing stages. Traditional translation evaluation would take place during the localization testing stage. Another study identifies linguistic, functional, visual and technical QA stages (Corte, 2000).
languages (Bass, 2006, p. 92). This is an essential aspect that distances localization evaluation from other translation evaluation processes and, given that translation evaluation has to incorporate contextual aspects (Sager, 1989), measurements have to account for both interrelated stages.

In the context of a time and resource limited process (Wright, 2006), the tension between technological and linguistic aspects of this type of evaluation has led the industry to place a higher emphasis on functionality testing, disregarding more complex linguistic aspects beyond the sentence or microtextual level of analysis (Dunne, 2006a)\textsuperscript{11}. In fact, out of eight componential evaluation areas in the LISA QA model, only one deals with "language", in which an error can be identified as a mistranslation, an accuracy error, terminology error, language error, style error, country or dialect error, and consistency error. Given that these errors do not have an empirical foundation; it is easy to perceive the difficulty in assigning any given error to only one of these categories. Some of them overlap, such as accuracy and style errors, and language and style errors. Moreover, consistency is the only category that could be assigned to some of the more serious inadequacies at macrotextual or superstructural levels, such as using different terms for the same section of a website. Thus, while these inadequacies are theoretically accepted to be more important (Larose, 1998; Williams, 2003), translators and evaluators are somewhat forced into a microtextual level that has difficulties in dealing with problems at the suprasegmental level, or in other words, those that affect the text as a global unitary entity. The same can be said of the more important pragmatic errors (Nord, 1997), as the

\textsuperscript{11} The researcher points out: "Because functional aspects of a program can be objectively assessed [...], people often perceive software development and localization as processes that are akin to manufacturing, while forgetting about the linguistic aspects of the program and the often subjective nature of linguistic quality definitions" (Dunne, 2006b: 96).
decontextualization of textual segments hinders the evaluation of a global coherent text from a functionalist and pragmatic perspective.

**The LISA QA Model and Functionalist Approaches to Evaluation**

Taking into consideration that this model responds to economic and time constraints, one of the positive aspects of the LISA QA model resides in the fact that it is componential, or in other words, that it recognizes many aspects of quality (Colina, 2008). This approach could be easily tailored to the notion of context dependent quality (Sager, 1989). It also implicitly accepts that the search for an absolute standard of quality is impossible. In this case and from a functionalist perspective, the localized text will be more or less appropriate for the purposes for which it is intended (Nord, 1997). However, the fact that the evaluator has to constantly decide on assigning weight to errors does not provide the sufficient degree of reliability and validity as stated by a number of researchers (Williams, 2003; Melis & Hurtado, 2001; Colina, 2008).

Additionally, this system follows a structural and componential approach based on the identification of a customizable number of language and transfer errors, even when empirical error typologies of errors have not been fully developed. Research in multiple language combinations could produce more reliable and efficient typologies that could incorporate recurrent pragmatic, functional and communicative inadequacies. These types of pragmatic and functional features are more difficult to evaluate than grammar, meaning or spelling (Colina, 2008, p. 125) and, therefore, they tend to be absent in error-based systems such as LISA’s. These pragmatic/publicist areas have been recently incorporated through holistic evaluation proposals (Waddington, 2001; Colina, 2008) that, among others, provides for a way of evaluating the communicative/pragmatic adequacy of a localized text, "rather than being limited
to grammatical and or stylistic errors at the sentence level, and changes in meaning" (Colina 2008, p. 107).\textsuperscript{12}

The holistic review evaluates the pragmatic/functionalist adequacy of the target text as a whole (Nord, 1997; Waddington, 2001), providing a more reliable assessment than what is possible through evaluating single segments alone. It uses descriptive statements that simplify the evaluation process, allowing raters to concentrate on specific inadequacies in the middle of a quality continuum, rather than correct/incorrect judgments. The statements used are attached to a numeric scale and they are implicitly formulated, such as the following example from Colina (2008):

1. Disregard for the goals, purpose, function and audience of the text. The text was translated without considering textual units, textual purpose, genre, audience, etc.

   […]

4. The translated texts accurately accomplishes the goals, purpose, (function: informative, expressive, persuasive) set for the translation and intended audience (including level of formality). (p. 129)

As mentioned previously, the development of an empirical error typology that would incorporate recurring pragmatic inadequacies, combined with an additional holistic level that could assess the entire text through descriptive statements, would evaluate more reliably whether texts look like they have been developed "in-country". In the development of these typologies and the holistic evaluation of the pragmatic adequacy, corpus based approaches to evaluation can provide a solid empirical foundation. The following section reviews these approaches, as well as proposes different practical ways to incorporate corpora into the evaluation of pragmatic and functionalist issues.

\textsuperscript{12} These methods used descriptive statements in order to describe the level of adequacy of any given translation to the intended function, genre, and text type of the target text. These holistic models have also been empirically proved to be as valid and reliable in summative evaluation processes as error based approaches (Waddington, 2001).
Corpus-based approaches to quality evaluation

Corpus-based approaches to translation research have been extremely productive during the last decade (Baker, 1995; Laviosa, 2002; Olohan, 2004). The success of this approach is due to the fact that, if the textual selection process is done according to established principles (Sinclair, 1991; Biber, 1993) it provides quantitative data that can support valid and objective decisions while translating or evaluating (Shreve, 2006, p. 311). This more objective and empirical approach to evaluation has been defended by numerous scholars (Nord, 2003; House, 2001; Melis & Hurtado, 2001; Bowker, 2001), and it is somewhat striking that little research on corpus-based evaluation in localization has been done so far. Among others, Bowker (2001, p. 346) defends the importance of using corpora in evaluation, since it "entails making judgments about appropriate language use, [and] it should not rely on intuition, anecdotal evidence or small samples". It could be argued that the same rationale would apply to localization evaluation, given that corpus approaches can provide the empirical data that can help overcome the subjective bias unavoidably present in all evaluation processes (Hönig, 1998, p. 14). Using corpora could make this process much more transparent, and it would allow evaluators to make decisions that can be easily backed up by representative quantitative data (Baker, 2004, p. 184). In this case and among different corpus types (Laviosa, 2002, p. 34-38), evaluation corpora can constitute a source of helpful conceptual and linguistic information (Bowker, 2001).

Few translation scholars have applied corpus-approaches to localization. Its practical use during professional translation and evaluation has been both supported (Shreve, 2006; Jiménez-Crespo, 2009, 2008a) and rejected (Biau & Pym, 2006). The main ground for rejecting the use of corpora is the fact that they are not cost-effective when applied to professional translation
While time and cost issues are central in professional summative evaluation, it is fair to defend that a bottom-up approach, compiling extensive comparable or translational corpora made up of original texts in the target language, and translated texts into that same language from English, could prove to be part of a more objective base for localization evaluation. The potential uses of these evaluation corpora are described in the next section.

**Uses of Corpora in Translation Evaluation**

The use of corpora in Translation research, training and professional practice has been widely explored during the last two decades (e.g. Laviosa, 2002; Olohan, 2004). Many of the established principles and procedures of this methodology can be highly beneficial to localization evaluation both in the preliminary stages and during the evaluation process. The following list briefly describes the main applications of this methodology:

1. **Descriptive monolingual and bilingual genre-based grammars or contrastive studies.**

As mentioned above, the industry has as an objective to release products that look like non-translated texts in any give locale (LISA, 2004, p. 11), and therefore, quality should be mostly related to the expectations of the receivers (Jiménez-Crespo, 2009; Nobs, 2006). Thus, the logical point of departure for corpus-based approaches to localization evaluation entails descriptive studies of conventions established in non-translated textual genres. This could be achieved by compiling evaluation corpora of the most recurring original and localized digital genres. In the case of bilingual corpora, contrastive analyses of original and localized sites can identify linguistic features that distance localized texts from the textual prototype shared by the target discourse community (i.e. superstructural differences, stylistic and terminological

---

13 English is the most frequent source language in localization (Cronin, 2003). Nevertheless, the Internet has multiplied the localization processes in the opposite direction, from any language into English, and between multiple languages in contact in specific areas, such as Italian, German, French, Rumantz and English in Switzerland.
variation, pragmatic differences, typology of recurrent translation and localization errors, etc.), and therefore, that could influence the receiver's quality judgment on the text. These descriptive studies can also shed some light on most functional and pragmatic inadequacies in localized texts.

2. Evaluation corpora (Bowker, 2001). The need for comparative empirical data in corpus studies is well known and documented (Stubbs, 1996; Kenny, 2001). Among different corpus types, a specialized control/evaluation corpus can be used as a benchmark against which the uses of any given term, collocation, phraseological unit, etc. can be quickly contrasted during a translation evaluation process. This is mostly needed in cases where the evaluator requires quantitative data in order to support or reject a possible decision. The process entails the use of a lexical analysis tool (i.e. Wordsmith tools, Paraconc or Monoconc) that provides applications such as concordancers, wordlists or keyword analysis. A concordancer can retrieve all instances of use of any given word or search pattern in its immediate context, and displays these in a clear format that can be reorganized according to different criteria.14 These concordancers allow wildcard searchers in order to identify and analyze in context different uses of a single word, such as "virus*" in order to retrieve all word forms for this lemma such as "virus" and "viruses". As an example, Figure 2 presents a concordance line for the noun "virus" both in the original and the localized corpus of the Spanish Web Evaluation Corpus. This search was carried out in order to observe the adequacy of a translated phraseological unit in a web privacy policy, "libre de virus" 'free of viruses':

14 A concordancer can be defined as a "tool that retrieves all the occurrences of a particular search pattern in its immediate contexts and displays these in an easy-to-read format" (Bowker, 2002: 53).
Pragmatic and Functionalist Aspects in Localization

Figure 2. Extract from a KWIC (Key Words in Context) display of the results of a query of the noun "virus" in original and localized texts into Spanish.

The data provided by the concordance analysis shows that the phraseological unit "libre de virus" does not appear in original texts in order to formulate this recurring unit of meaning. It can be thus recognized as a calque of the original English collocation "free of viruses", and therefore, evaluators can quickly establish that this is a pragmatically inadequate expression in the target locale. As a solution, in original Spanish texts this noun is used mostly in combination with the words "presencia" 'presence' or "ausencia" 'absence'.

3. Word frequency lists. Frequency lists can be used in order to identify possible patterns of error during the translation evaluation process. These lists allow evaluators to discover how many words there are in a corpus (tokens) and how often each of them appears (types), thus providing evidence about recurring lexical or syntactic patterns of interference. Lexical analysis tools such as Wordsmith Tools can provide lists not only for single words, but they can also provide lists of word clusters that co-occur in a given corpus. These lists can also help spot recurring terms or phraseological units that, due to their high frequency, can be identified as borrowings or calques, such as "términos y condiciones" 'terms and conditions' or "Acerca de"
nosotros" 'About us' in localized websites into Spanish (Jiménez-Crespo, 2008a). Frequency lists for single words or clusters of words in a given text can also be compared with a tool such as Keywords with the frequency list obtained in the original or control section of the evaluation corpus. These tools compare the frequency of a single word or cluster of words in terms of percentage of use in the overall corpus, and once the frequency list of a text to evaluate is contrasted with the list of the original evaluation or control corpus, the list obtained can show recurrent lexical and syntactic patterns that are overrepresented in contrast to original texts. This methodology can help spot cases of lexical or syntactic interference in localized texts. As an example, Table 1 consists of a keyword analysis of localized legal texts in websites using the Spanish original websites as a control corpus. Once proper names were eliminated, the analysis shows that the frequency of the pronoun "usted" 'you' is six times higher in localized texts. This is due to the fact that pronouns that function as subjects in Spanish can be omitted, while they are compulsory in English. Additionally, the word "leyes" [laws] is also more frequent in localized texts. This higher frequency can be traced back to the fact that in the Spanish legal texts the concept "laws" is expressed by the term "legislación", thus the higher frequency of the word "leyes".
Table 1.

*Keyword Analysis of Localized Legal Web Texts in the Evaluation Corpus Using the Subcorpus of Original Legal Web Texts as a Control Corpus. Names of Corporate Businesses were Eliminated.*

<table>
<thead>
<tr>
<th>N</th>
<th>word</th>
<th>freq.locExt %</th>
<th>freq.netLeg-1.txt %</th>
<th>keyness</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>usted</td>
<td>433</td>
<td>36</td>
<td>0.06</td>
</tr>
<tr>
<td>3</td>
<td>sitio</td>
<td>1,212</td>
<td>283</td>
<td>0.49</td>
</tr>
<tr>
<td>8</td>
<td>información</td>
<td>1,195</td>
<td>327</td>
<td>0.57</td>
</tr>
<tr>
<td>13</td>
<td>privacidad</td>
<td>317</td>
<td>61</td>
<td>0.11</td>
</tr>
<tr>
<td>17</td>
<td>terminos</td>
<td>251</td>
<td>46</td>
<td>0.08</td>
</tr>
<tr>
<td>22</td>
<td>linea</td>
<td>55</td>
<td>0</td>
<td>45.6</td>
</tr>
<tr>
<td>32</td>
<td>declaración</td>
<td>85</td>
<td>0</td>
<td>42.7</td>
</tr>
<tr>
<td>41</td>
<td>marcas</td>
<td>202</td>
<td>41</td>
<td>0.07</td>
</tr>
<tr>
<td>42</td>
<td>utilizar</td>
<td>214</td>
<td>45</td>
<td>0.08</td>
</tr>
<tr>
<td>43</td>
<td>leyes</td>
<td>108</td>
<td>12</td>
<td>0.02</td>
</tr>
</tbody>
</table>

This same type of analysis can be performed with word clusters of different length, normally up to eight words, and this could provide evidence of recurring syntactic, stylistic or lexical constructions that might be inappropriate when contrasted to the control corpus of original texts.

4. **Identifying patterns of pragmatic error.** Quantitative analysis of the translational component of the corpus could be useful in establishing and identifying patterns of errors, as well as classifying them (Jiménez-Crespo, 2008b; López & Tercedor, 2008; Melis & Hurtado, 2001). As previously mentioned a classification could help establish a more reliable objective foundation during error judgments in general or while assigning a certain weight to any identified error. Additionally, a classification of the most recurring errors can direct the attention of evaluators towards more serious error patterns, such as pragmatic and superstructural inadequacies that are more difficult to identify and, therefore, require higher levels of competence or specific training from the evaluators (Williams, 2003). As an example, the analysis previously presented with the verb "leer" 'to read' in the translational component of the
Evaluation Corpus can be revealing. In Figure 3, the concordance lines associated with the verb "leer" in the localized corpus show differences in its grammatical use and in its patterns of lexical co-occurrence if compared to the original Spanish texts in Figure 1.

Figure 3. An extract from a KWIC display of the results of a query of the verb "leer" 'read' in localized texts. All inadequacies or pragmatic errors are underlined.

All items underlined can be considered language errors, pragmatic inadequacies or non-conventional items in original Spanish texts. For example, the first line shows a translation of the lexical unit "privacy policy" that has been rendered as "intimacy policy". These concordance lines also show the terminological variation in the translation of "privacy policy", "política de intimidad", "política de privacidad", "normativa de privacidad", "política de confidencialidad". Furthermore, it can be observed that, while the direct appeal to the user is rarely present in original Spanish legal web texts and he/she is addressed impersonally (see Figure 1), in most localized concordance lines the user is addressed directly: "lea cuidadosamente…” ‘read carefully…’, "le recomendamos que lea…” ‘we recommend you to read…’, "le agradecemos que lea…” ‘we would appreciate it if you would read…’, "asegúrate de leer…” ‘make sure you read…’.
Moreover, a contrastive genre-based study can link all identified errors or inadequacies to possible solution(s) based on the range of natural possibilities that appear in original texts. For example, the translation of the lexical unit "case studies" or "success stories" used in navigation menus in corporate websites could be linked to those lexical units that serve a similar function in original texts, such as "experiencia de la empresa" 'Company experience' or "nuestros clientes" 'our clients' (Jiménez-Crespo, 2008a, p. 369).

Conclusions

In interactive texts such as web pages or software products, the pragmatic adequacy of the localized texts is directly linked to the success of the communicative interaction established between the digital text and the receiver. Once established that the goal of the localization industry for localized texts is to look like "locally-made products" (LISA, 2003, p. 5), this paper has defended that QA systems should go beyond structural approaches centered on the identification of lexical, syntactic and transfer errors and incorporate an additional pragmatic level. Pragmatic inadequacies are harder to identify by native speakers than lexical or grammar errors: they are not erroneous per se in the target language system, but more or less inadequate with regard to the communicative purposes intended (Nord, 1997). In order to add this additional level, this paper has highlighted the importance of providing a theoretical and methodological base in order to successfully asses this integral aspect of translation evaluation.

As previously mentioned, the scope of this paper was not to propose a novel evaluation model or apply existing complex theoretical proposals, but rather, to set in place this much needed theoretical and methodological foundation as a first step towards a more valid and
reliable QA model. Given that localization is an industrial process with limited time and resources (Sager, 1989, p. 91), the contribution of this paper can be summarized in a proposed triple approach that can complement current existing models: (1) the identification of pragmatic errors through carefully constructed representative corpus studies in order to include them in QA error-based systems (Jiménez-Crespo, 2008a), (2) the use of evaluation corpora during the assessment processes (Bowker, 2001), (3) the possible addition of a global holistic system as an additional step that incorporates a module for the pragmatic and functionalist adequacy of the target text (Colina, 2008; Waddington, 2001).

Finally, this paper has highlighted the need for more theoretical and applied research into localization in general and into evaluation in particular, as the fuzzy definitions of operative notions, such as "quality", "error", "language error", "error typology", "reliability" or "evaluation objectives", hinder to some extent the development of solid QA systems that can produce more objective valid and reliable outcomes. It is hoped that this paper will provide a starting point for establishing an agreed-upon evaluation model that can benefit not only practitioners and researchers, but also the end-users of localized digital texts.
Pragmatic and Functionalist Aspects in Localization

References


Miguel A. Jiménez-Crespo

Martín (Eds.), *Aproximaciones cognitivas al estudio de la traducción e interpretación* (pp. 259-300). Granada: Comares.


Pragmatic and Functionalist Aspects in Localization


Miguel A. Jiménez-Crespo


