Localization and writing for a new medium: a review of digital style guides

Miguel A. Jiménez-Crespo
Rutgers University, USA

Abstract
Over the last two decades, an increasing number of publications have described the differences between print and screen writing, both from prescriptive and empirical perspectives. These recommendations are of paramount importance to localizers, as the ability to adjust their writing style to different mediums, genres and communicative situations is one of the essential skills these professionals should acquire. This paper offers a review of recommendations in web and software writing style guides in order to provide a concise summary for localization professionals and trainees. The overview encompasses guidelines from language, localization type, company-specific style guides, and usability publications.

Keywords
localization, style guides, technical writing, digital texts, multimedia translation

1. Introduction
Localization is rapidly becoming one of the sectors with the highest volumes of translation and an increasing number of professionals are becoming experts in software, videogame or web localization (Mossop 2006). Among the many technological, communicative, textual, and pragmatic challenges that localizers face, adapting their writing style to the demands of screen reading is an essential skill that needs to be acquired. Usability studies have shown that reading on screen is normally 25% slower than on paper (Nielsen and Loranger 2006) and, in general, users do not "read" but scan through screen texts until they find what they are looking for: simply what interests them (Nielsen 2000). The interactive nature of screen texts demands a specific writing style in order to reproduce the texts with the highest levels of usability, a variable directly related to the success of any digital product. Localizers are directly responsible for text production, and therefore, a sound knowledge of writing styles for on-screen reading is part of what has been called "professional localization competence".

Additionally, in a process such as localization where most processes are normally performed by teams, the lack of a consistent style is considered a common issue in localized products. Localized texts with heterogeneous styles are what researchers refer to as "sentence salad" (Bedard 2000), "train wrecks" (Bowker 2006) or "collage texts" (Mossop 2006). As a result, these texts can be inherently less cohesive or coherent, less readable, with oversimplified syntax and a monotonous rhythm (Bowker 2006: 180). Style guides play an essential role in establishing a more coherent and cohesive style. They have been defined in localization as an aid to "create documents appropriate for the end-user, for meeting company and country standards, and for maintaining geographic and cultural suitability" (Lingo Systems 2004: 29). A review of all style guides specifically developed for digital texts and/or localization shows four distinct categories: (1) style guides for specific localization types, such as those that apply to web localization.
software, (2) language-specific style guides, (3) usability style guides, and (4) company and project-specific style guides. Moreover, styles guides can be subdivided between those that deal with digital style in general, and those that focus specifically on interlinguistic localization processes, such as the Microsoft Language Portal (http://www.microsoft.com/Language/en-US/Default.aspx). For this category, European Translation Quality Standard 15038 indicates that they should include specific information about punctuation, spelling, formatting, adaptations, language-specific and client preferences, common errors to be avoided, and other miscellaneous elements.

In order to provide a concise review for localization professionals and trainees, this paper summarizes the recommendations from the most representative style and development guides. The publications used for this comprehensive review are:

1. Language-specific web and software style and development guides:
   b. In Spanish (Díaz Noci and Salaverría 2003; Vilamor 2001)
2. Company-specific style guides: Google, Microsoft, Oracle-Sun or Apple
3. Discipline-specific guides, such as web and software localization publications (Esselink 2000; Yunker 2003; 4; Chandler 2004; Lingo Systems 2004: 60-66)

Recommendations in these style guides are generic in nature. It should be noted that they do not necessarily apply equally to all textual elements to localize (Gregory 2004), and additionally, the fact that style guides exist does not mean that translators will necessarily follow their recommendations (Gamero 2001). Digital genres inherently include a varied mix of textual elements with different communicative purposes. For example, websites include both "content" text, the distinct content in each page, and "interface" text, all the textual elements that help structure the global hypertext: such as navigation menus, banners, site maps, search functions, pop-up messages, etc (Price and Price 2002). Each hypertextual type entails a differentiated communicative situation: in "content" texts, a company might be directly communicating with end-users, while in "interface" texts; the communication process is established between the website and the end-user (Janoschka 2003). Similarly, the multiple conventionalized sections in each website, such as contact pages, "about us" pages or legal disclaimers might require different writing styles from the localizer (Jiménez-Crespo 2008). Thus, these generic recommendations should be viewed as a compendium of differences between print and digital writing styles. In any case, it is clear that producing high-quality usable, credible digital text is one of the many professional skills any localizer must possess. Adjusting target texts to the expectations already shared by target users requires special attention to recommendations from technical writing, development, and usability guidelines. For convenience of use, this review has been divided into general guidelines: semantic-lexical aspect, syntax, stylistic aspects, typography, iconic-visual aspects, and editing-reviewing.

2. General guidelines

All publications contain the same basic recommendations: quality digital texts must be brief, concise, informal, clear, accessible, and specific. These basic recommendations are slightly comparable to the basic tenets of scientific and technical writing (Duque 2000:21-22): clarity, conciseness, precision, and objectivity.

- Briefness is related to the peculiarities of screen reading, specially the fact that users scan screens until they find the content that interests them (Nielsen and Loranger 2006). This reduces the level of noise in the text (Krug 2005), consequently decreasing the processing cognitive load and highlighting the specific
content in each section or page. Normally, the consensus is that web texts should be up to 50% shorter than similar printed texts (Nielsen and Loranger 2006). In order to achieve this goal, unnecessary repetitions and content that is common knowledge among the targeted discourse community should be avoided (Price and Price 2002:88). Localizers should pay attention to unnecessary words such as really, just, etc. as well as demonstrative adjectives and pronouns. The search engine Google is usually presented as the optimal example of brevity and conciseness, hence part of its success (McAlpine 2001).

- Conciseness is as important as briefness. It entails one of the challenges of web writing (Mill 2005). Conciseness is directly related to lower cost for later localization efforts (Lingo Systems 2004).
- Clarity is essential in all web texts. Users have less tolerance for ambiguity in interactive screen texts than in printed texts (Bly 2002).
- For English texts, non-native speakers should be borne in mind while translating (Kilian 1999; McAlpine 2001; Lynch and Horton 2009). Furthermore, including culture-specific play on words, metaphors, Americanisms, etc. is not recommended. (McAlpine 2001: 45; Yunker 2003)

2.1. Lexical-semantic aspects

- Use simple terms that are familiar and easy to understand for users, avoiding sophisticated or low frequency terms (Nielsen and Loranger 2006; McAlpine 2001; Price and Price 2002; etc.).
- Be coherent and consistent in the terminology used, avoiding several synonyms for the same concept (Yunker 2003; Jeney 2007; Bly 2002; Price and Price 2002).
- Avoid using acronyms and abbreviations. When used, they should be spelled out (Mill 2005), however it is possible to use widely used acronyms such as DVD, PC, etc. (Nielsen and Tahir 2002). In any case, they are appropriate for websites or sections addressed at specialized or semispecialized users (Garrand 2001). Similarly, technical terms should be avoided if the section is not addressed at specialists.
- Avoid slang or jargon (Nielsen and Loranger 2006: 262), clichés, or generalizations (Jeney 2007).
- Sarcasm and humor should be used sparingly, and whenever used, it should be extremely clear (Mill 2005: 26). These discursive devices can lead to comprehension problems and confusion (McAlpine 2002: 41). The use of both in navigation menus should be avoided at all costs.
- Attention should be paid to dialectal variation and regional uses as sites can be accessed from any part of the world (Kilian 1999; McAlpine 2001: 45).
- In English, it is preferable to use words of Anglo-Saxon origin over Latin, such as "get-obtain", "end-terminate" (Price and Price 2002; Bly 2002).

2.2. Syntax

- Syntax should be simple, avoiding long concatenated sentences (Horton 1990).
- Sentences should be simple, short, and complete; with one idea per sentence. In English, it is recommended that sentences should have a maximum of 15 to 25 words (Mill 2005; Yunker 2003). This might be different for each language. For example, sentences in technical Spanish are, on average, longer than those in English.
- Use the active voice (McAlpine 2001: 47; Jeney 2007). Users find the active voice easier to process (Bly 2002).
- Users scan textual sections in the first stage (Nielsen 2000; Nielsen and Loranger 2006). Paragraphs should start with short sentences that contain the main idea, a basic principle of the inverted pyramid structure. This structure starts with the conclusion, then offers the important information, and ends with the rest of the situational information (McAlpine 2001: 44; Nielsen and Loranger 2006: 269). According to Mill, "the effectiveness of the inverted pyramid style of writing on the Web is because the reader […] can quickly consume the key information at the beginning" (2005: 11-12).
- Use verbal forms instead of noun forms when possible (Price and Price 2002). Avoid using too many phrasal verbs in English. They are more difficult to understand for non-native speakers of English (McAlpine 2001: 48).
- Using single verbal forms over periphrasis, such as "decide" for "make a decision" or "use" for "make use of" is also recommended. The same applies to prepositions and conjunctions, such as using "although" for "despite the fact" or "but" for "notwithstanding" (Mill 2005).
- Use descriptive nouns and adjectives (Garrand 2001). Use simple verb tenses when possible (McAlpine 2001: 48).
- Use caution when using anaphoric, cataphoric, and deictic pronouns. It is impossible to predict whether the user will read the rest of the hypertext (Price and Price 2002). The same applies to endophoric, exophoric and deictic references, demonstrative pronouns, etc.

2.3. Stylistic level

- A conversational tone, similar to oral speech, should be used (Mill 2002: 26). The text should address the user directly (Bly 2002: 143; Nielsen and Tahir 2006: 262). This recommendation appears in most publications: "The Internet works best as a series of two way conversations" (Price and Price 2002: 31), "[...] write as if you are talking to an individual, not a collective group of anonymous web surfers" (Garrand 2001: 32), "use a friendly, lively, individual tone. Remember, you are conversing with one person at a time" (McAlpine 2001: 37). This does not mean that this text-type convention necessarily applies to other languages. For example, Spanish websites favor a more impersonal tone than English ones (Jiménez-Crespo 2008).
- Include a positive attitude (Mill 2005). This can be achieved by excluding double negatives and, instead, using positive expressions such as "little" for "not much" or "different" for "not the same". When using verbs, combining negative expressions with verbs carrying a negative semantic load, such as "avoid, deny, exclude, lack, prevent, prohibit", or with prepositions such as "against" or "without" is not recommended (Price and Price 2002). This is recommended given that negative expressions also require a higher comprehension cognitive effort (Sammons 1999).
- The use of an informative style is most effective on the Internet. Usability research has shown that the informative style is 25% more effective than the marketing-advertising style (Nielsen 1999) since it increases users’ readability and recall capacity. According to Bly (2002:12) "write your copy like you are sharing information, rather than selling".
- Avoid stereotyped language such as "site under construction", "welcome to my site", "click here!" or "check this out!" (Kilian 1999: 11; Yunker 2003; Nielsen and Loranger 2006). These expressions have also been calqued and incorporated into most languages.
- Use appropriate punctuation. Avoid semicolons, in some cases, as they might not be identified in screen readers (Price and Price 2002). In case of doubt, split sentences into two.
- Dividing the information into two short paragraphs helps the scanning process. Each paragraph should have one idea or communicative purpose. For English, a paragraph should have from 3 to 5 sentences (Mill 2005; Price and Price 2002: 91), with 100 words at most.

2.4. Structural level

- The text should be chunked if it is longer than two screen lengths. Nevertheless, each page should be clear and be a closed unit by itself (McAlpine 2001: 13). The most important information should be presented in the first two sentences (Nielsen and Loranger 2006) and each page, as a unit of content and storage, should be described in a single sentence (McAlpine 2001: 26). Text chunking should not be done arbitrarily (Lynch and Horton 2009). One exception to the chunking rule is a long piece of news, as it can be divided into two or three pages (Price and Price 2002: 147).
- Repetitive or numeric information should be included in tables and graphics. This improves the identification of relevant information for users (Bly 2002: 73-78).
- If a paragraph should have four or more elements, using lists is advised (McAlpine 2001: 5), even though this strategy should not be used often (Nielsen and Loranger 2006: 280). Listings should not begin with a demonstrative.
- Use descriptive, concise titles and headings, as they can help capture the user’s attention (Mill 2005: 3). Titles should have less than sixty characters (Bly 2002). Eye tracking studies have shown that users only scan the first words in titles. Attention should therefore be paid to the initial words in titles and headings.
- Hyperlinks structure the entire hypertext. For that reason, the terms associated with them should be carefully localized (McAlpine 2001: 33). These hyperlinks are conventionally marked by underscore and a different font color and hence should not be implicitly signaled with redundant expressions such as "click here!" or "Follow this link to..." (Price and Price 2002; Mill 2005: 44).

2.5. Typographic aspects

- Using bold for emphasis is preferred to italics, because italics are harder to read on screen (Nielsen and Loranger 2006). Underlining is not recommended as this is the conventional marking for hyperlinks (McAlpine 2001: 61; Nilsen 2000). Using all caps for emphasis is not recommended because it slows down reading by 10% (Nielsen and Loranger 2002: 238).
- Use sans-serif fonts designed for screen reading, such as Verdana, Georgia, New York and Trebuchet. Usability studies have shown that these fonts improve screen reading (Nielsen 2000).

![Serifs vs No Serifs](image)

Figure 2 Differences between sans-serif and serif fonts.

- The optimal font size is 12, although size 10 is also possible (Nielsen and Loranger 2006). Nevertheless, it should be noted that users can adapt the font size to their preferences or screen resolution.
- Do not use more than four colors as a whole (Nielsen and Loranger 2006: 235). Black on white is the most legible combination, followed by blue on white. Hyperlinking conventions should be avoided when using font colors: magenta for visited links and blue for unvisited links (McGovern et al. 2002).

2.6. Iconic-visual aspects

By nature, multimedia digital texts are a combination of textual and iconic elements (Tercedor 2005). The metaphorical use of images has become an essential aspect of multimedia communication (Posteguillo 2003: 63; Janoschka 2003). However, all style guides recommend the minimal use of images. As Jeney (2007) rightly mentioned:

Web writers are frequently tempted to replace textual information with images, but it is important to remember that images are usually static and nonhierarchical in nature, whereas the human mind prefers structure and hierarchy [...] Only use the images so that they fit the purpose of the website (Jeney 2007: 28-29).

The following recommendations help to understand the interrelated nature of text and image and how best to deal with them.

- Images should only be used whenever they complement the communicative purpose of the webpage. Images should not replace textual elements as they require interpretation (Jeney 2007:28-29).
- Eye tracking usability studies by the Poynter.org Institute (http://eyetrack.poynter.org/) confirmed that users tend to concentrate on texts instead of images (Nielsen and Loranger 2006).
- Despite the enormous contribution of images to the communicative purpose of the site, text is the main element. According to Kilian (1999: 15) "The graphic elements of a web site are especially important to them [users], but text remains the core of most websites".
- Images in texts cannot be searched for using search engines (Nielsen and Loranger 2006: 7). In case there is text embedded in the images, it should be included in the content of the html attribute "alt". It should be mentioned that embedding text in images is not recommended for texts to be localized (Lingo Systems 2004: 60).
- All style guides encourage including the descriptions of all ATL attributes in images. These attributes are of utmost importance for disabled users and are required for publicly financed websites in many countries (Tercedor and Jiménez-Crespo 2008). These descriptions need to be clear and concise for any image or icon, especially if they serve as a navigation link. McAlpine (2001: 101-113) includes the following summary to describe images and icons in websites.
  
  a. Before describing the image, it is essential to know the function or purpose of the image.
  b. Leave the alt attribute empty for decorative images only or for those used to organize the text alt="". If the empty "alt" attribute is not included in these cases, screen readers for disabled users will repeat the word "image" each time it encounters an empty attribute.
  c. If the image is used to end a paragraph or list, write "end of section", "end of page", etc.
  d. All alt texts should end with a period, otherwise the screen reader will link this text to the following segment without a pause, e.g. alt="Company logo."
  e. If an image is part of the content, such as graphics, tables, or images with embedded text, describe all information in its entirety.
  f. Quality in these segments is essential given that search engines and indexers use them to describe and index the pages.

2.7. Text editing/QA analysis

In general, all publications agree that editing guidelines are somewhat relaxed for web texts (Posteguillo 2003). Nevertheless, a top quality analysis stage is paramount. The following recommendations were found:

- It is required to perform a complete quality editing stage (Bly 2002: 56).
Errors are associated with a lack of professionalism or quality on behalf of the company or organization behind the text. "In a commercial context, writing errors are seen as evidence of weak and sloppy thinking. This, in turn, reflects upon the company or institution that owns the site" (Jeney 2007: 46). Errors can significantly decrease the credibility of any website. Relying exclusively on automatic spelling and grammar correctors is not recommended (Nielsen and Loranger 2006; Jeney 2007). The following is a brief summary of the most common errors in localized digital texts (Corte 2002; Jiménez-Crespo 2008: 536-560; Jeney 2007):

a. Spelling, grammar, and typographic errors
b. Wrong terminology
c. Inconsistent terminology, typographic uses, tone to address the user, etc.
d. Incongruent text-image relationships
e. Source language segments in frames or pop-up messages
f. Commonly mistaken words in English, such as accept-except and conscience-conscious.
g. Lexical, syntactic, and typographic anglicisms
h. Ambiguous or unclear sentences as a result of translation errors

3. Conclusions

Adapting writing styles for on-screen reading is one of the many skills professional localizers possess. It is hoped that this review will be useful for localization practitioners, trainers and trainees alike. From a functionalist perspective (Nord 1997), the localization profession entails loyalty to both commissioners-initiators and to end-users. For localizers, producing the most usable, effective target texts means not only fully meeting the clients’ expectations, but also improving end-users’ experiences while interacting with localization products.

References


Anex I. Links to style guides

2. Oracle Style guides: http://wikis.sun.com/display/g11nhome/Language+Style+Guides