INTERNATIONALIZATION AND TRANSLATABILITY

David Kumhyr, Carla Merrill, Karin Spalink

Internationalization and Translation Services, Inc.
3503 Stonegate Drive
Durham, NC 27705-5480
tel: 919-490-4963
fax: 919-490-2069
email: 70720.1406@compuserve.com

Abstract

Marketing to a global clientele requires adaptation to the clientele's locale. Adaptation can be built into the production process or it can be achieved through more costly and time-consuming retrofitting. Local standards, cultural aspects and the meaning of symbols and icons have to be observed and incorporated; in other words, production for a global marketplace should be done with translation in mind. Translation considerations vary depending on the translation means to be employed.

1 Introduction

What is internationalization and who needs it?

Internationalization, sometimes confused with localization, has not yet become a household word. It still has the aura of the far-away and maybe even the exotic. It sounds like something that everybody else needs, not us. In fact that is exactly the point, however, we are the ones who have to provide it, if we want to be viable players in the global marketplace. More and more, companies cater to more than one locale, and percentages of revenue from sales in foreign countries are growing from year to year [Taylor 92]. "The globalization of economic activity is perhaps the defining trend of our time" [O'Hara-Devereaux & Johansen 94]. To profit from participation in the global marketplace, a product has to meet the requirements of the selected target markets. And to maximize those profits, the product has to be prepared and designed to facilitate adaptation to the various markets. This preparation is called internationalization or enabling. The adaptation to the various individual markets or locales is called localization. The localization process itself is independent of the state of internationalization of a product. The costs that are incurred during localization of products that have not been internationalized (called retrofitting) are typically many times higher than the costs that are incurred for localization of enabled products. The question of internationalizing is, in essence, not a matter of convenience or aesthetics but one of economics [Hall 93]. What is translatability and how is it used? Translatability is usually expressed as an index or factor to represent the complexity of a given text. Text complexity may be defined for a particular environment, for example, a language direction or a machine translation system. Since it is a relative measurement no absolute measurement parameters can be given. Some of the most common parameters used are: sentence length, nested clauses, and homographs [Gdaniec 93]. If

the text is measured for translation by a machine translation system, text-external factors like tables, graphics, and file format may be included in the accounting of translatability also. The translatability index indicates the text's suitability for translation processing. It may be used to estimate the time and effort (cost) it will take to translate the document.

2 Internationalization Factors

Product internationalization is not a process in itself but it is one aspect that has to be considered at each step of the production process, starting with product design and ending with product support. A product can be anything from an airplane to an accounting software package, and from an instruction sheet on how to tie a bow-tie to a marketing brochure for your consulting service. The internationalization factors vary from product to product. A complex product like an airplane incorporates all the internationalization factors of all the product subsets, e.g., electronic devices, instruments, seating, software products, directional signs, maintenance guidelines, and so on. Internationalization of a simpler product like a bow-tie instruction booklet involves far less. Internationalization factors fall into two categories: product-independent and product-dependent factors. The following paragraphs contain some selected examples for each of these two categories.

2.1 Product-Independent Factors

Product-independent internationalization factors apply in all cases; they are not tied to the dimensions, performance or other features of the product.

2.1.1 Religious Aspects

Insensitivity towards religious preferences can jeopardize a product for an entire market. Just a few months ago McDonalds Corporation offered bags to their customers that were imprinted with flags from all over the world. Among others the bags showed the Saudi Arabian flag which includes a verse from the Koran. That these bags were used to hold greasy products and would subsequently be crushed and thrown into the garbage, was offensive to members of the Islamic faith. The bags therefore had to be removed from the market, an expense that probably could have been avoided with proper internationalization. Flags are subject to different rules and regulations in different countries, and their representative values are viewed and felt differently by different peoples.

2.2 Cultural Aspects

Information storage is another example that is laden with history and which has evolved differently in different cultures. Scrolls are probably no longer considered suitable for the storage of today's information in any country. The filing cabinet that has served that purpose in the United States and in most of the western European countries for so long is by no means universal. Modern information storage media like external hard drives, CD-ROMs, and the like are still far from mainstream in many countries of the world. If your product uses icons to indicate information storage media, it is very important to be aware that a symbol like the filing cabinet is not transferable to just any

target locale. Other symbols that need to be scrutinized in the internationalization process are holiday symbols (Christmas Tree, Thanksgiving Turkey, etc.) and landmarks (the Capitol, Mount Rushmore, Hoover Dam, etc.).

Just like symbols, formats are often not transferable, e.g., the US date format MMDDYY and the European date format DDMMYY. This difference is important to know if you are dealing, for example, with contracts that include a delivery date. According to the US format, a delivery date 061295 will be June 12, 1995; according to European interpretation, this means a delivery date of December 6, 1995. Other formats that fall into the same category are currencies, some units of measure, etc.

Products like installation documentation may include examples that use proper names like "John" or "Mary". In many cultures it is very impolite to call anybody by their first name unless they are family or very good friends. An internationalized product will use other means to transmit the intended content.

The approach to conceptual analysis differs among cultures. Some cultures are used to a top-down approach versus a bottom-up approach, a deductive versus an inductive method, and/or a left-to-right approach versus a right-to-left approach. These differences need to be reflected in the use of symbols and formats, documentation and navigation in general [Horton 93].

2.3 Product-Dependent Factors

Product-dependent internationalization factors pertain to specific features that the product needs to comply with, for example, toxicity regulation and transport dimensions, or that will be acceptable to the market locale, for example, colors or certain types of ingredients and so on.

2.3.1 Manufactured Goods

Internationalization includes the verification of norms and standards and their applicability for the international market. If for example the transport of your product is subject to specific transportation rules and regulations of the US Department of Transportation, this does not mean that your potential target markets will have the same or similar rules; looser or stricter regulations may apply. If such potential pitfalls are pointed out at the time of product design, potentially costly product changes can be avoided.

2.3.2 Software Products

Text expansion due to translation is probably the best known internationalization factor for software products. Text expansion is most visible in messages and pop-up windows. Input masks also cause problems, for example, when a zip code field will only accept numeric characters. It is unacceptable to customers that the translated version will have to resort to often indecipherable abbreviations because the programmer failed to construct the program by using dynamic space allocation features.

A popular method to avoid programming repetitive message elements is to compose messages from a set of static elements. Unfortunately this method (which seems very efficient for the single-language market) is detrimental to internationalization. Many languages require changes in adjective, noun, and/or verb endings depending on gender and/or number. Composing messages could

lead to grammatically incorrect renditions. Internationalization requires message text to be isolated as complete text into resource files.

A step up in complexity is the incorporation of bidirectional language flow and multi-byte character sets. There are primarily two market groups that require program adjustments with regard to character size and direction: the Arabic and the Oriental markets. Each of these environments has its particular challenges, e.g., in Arabic characters change depending upon the adjacent characters, and the same sentence may contain right-to-left and left-to-right elements. For the Oriental languages, character input presents a tremendous problem. One promising approach to facilitate the information interchange is Unicode.

Collating sequences are determined by the character set of the particular language. Because not all languages use the same character set, the collating orders are different, too. It is therefore important that collating sequences are not hard-coded in the original source code [ADAPSO 91] By the same token, the ordering of numbers, symbols, and words with initial capital letters in alphabetical lists is also handled differently in various locales [Hall 94]. The program has to be flexible with regard to these factors.

2.4 Documentation

Most of the documentation that is destined for foreign markets today has been produced electronically. It contains not only text but also graphics and sometimes even moving pictures and sound The document setup has to take into account that text can expand during translation by up to 50%. The document has to accommodate such expansion without causing any mismatches between frames and corresponding text in tables or lists. Text expansion should not cause any text boxes in a graphic to overflow and/or reach into the next box. Page-numbering has to be linked to the structure of the documentation, for example a chapter, and should not be fixed.

3 Translatability Factor

Translatability is a measurement of the time and effort it takes to translate a text. It is based on the complexity of the text, as described through parameters (like sentence length, homographs, abbreviations), and the breadth or limitation of features of the translation means. For a text in which the complexity parameters match the features of the translation means, pre- and post-editing will be minimal, keeping costs (time and effort) down. Translatability assessment can only take place after the writing process is completed. However, if document production is done under strict guidelines that have been established especially to achieve a high degree of translatability, a positive outcome can be predicted. Depending upon the environment, particular parameters might have to be considered to determine the suitability for a specific means of translation.

3.1 Generic Translatability Parameters

The following paragraphs describe some parameters that will have to be considered in every translatability assessment, independent of the environment.

3.1.1 Sentence Length

Sentence length and complexity generally go hand-in-hand, the longer the sentence, the more complex it is, i.e., the more difficult it is to understand. Sentence length increases are usually caused by adding clauses which compounds the complexity of the sentence structure. A sentence that is difficult to understand carries a high risk of misinterpretation and misrepresentation in the target language.

3.1.2 Homographs

The success rate for homograph disambiguation is closely related to the amount of context that is provided. A human translator not only looks at the context immediately surrounding the homograph but at the text as a whole and at the same time brings his or her world knowledge into play. The machine translation systems of today, however, are not able to draw on the context or world knowledge as a resource for homograph disambiguation. Homograph disambiguation to them presents one of the major hurdles. Homographs can appear in short sentences as well as in long sentences. A single ambiguous word carries the highest risk of misinterpretation. Because of the increased risk for misinterpretation from long complex sentences, the general rule for writers is to write short sentences and if possible not to use any homographs.

3.1.3 Abbreviations

Most abbreviations fall into one of the following categories: mnemonics, conceptual equivalents and conceptual inequivalencies.

Mnemonics: These types of abbreviations are often found on display panels. If they are not translated into the target language, the user has to first learn the original source word and then its abbreviation. It would therefore be much more desirable to translate the mnemonic into a sensible abbreviation for the target term. A mnemonic is usually used because of space restrained. Frequently the first three letters of the English source word is used as the mnemonic. In languages that use a lot of prefixes, the first three letters are seldom usable to distinctly differentiate one word from another. Meaningful translation of mnemonics therefore requires point-of-translation decision making which up to now can only be achieved by human translators.

Conceptual Equivalents: Abbreviations for conceptual equivalents also create a problem for machine translation. Example: Use "O" for Operator. The concept of "operator" also exists in the target language, however, the target word does not start with "O". Establishing a connection between the message of the sentence and the required change in letters is not immediately possible for a machine translation system.

Conceptual Inequivalences: Abbreviations for concepts that do not exist in the target locale usually require an explanation rather than a translation. Such an explanation can be given by a machine translation system just as well as by a human translator. It could presumably be made either the target dictionary entry for that abbreviation or it could be made part of that entry.

4 Translation/Localization

Production for a global marketplace, as we have said before, has to be done with translation in mind. Internationalization prepares a document for translation. It minimizes ambiguities, simplifies the text, and isolates it from code and markup language. Internationalization not only facilitates manual translation but also prepares your document for machine or machine-aided translation. The language restrictions imposed by internationalization guidelines at the same time restrict the number of possible mistranslations and misconstructions by the translation software. There is a dependency between document internationalization and document translatability.

If it has been established at the outset of the production process (as it should) that the document will be translated by machine, the list of internationalization factors has to include machine system specific features. This can be as simple as checking for format capabilities, e.g., if your documentation contains version or part numbers that include periods, does the system recognize those as such or does it interpret them as end of sentence markers? It could also be as complex as discovering that the system can not process the file format of the software package that you intended to use for the production of the documentation. In this case it might be necessary to switch software packages or to develop a file format filter or conversion program.

A fully internationalized product is in general easier to translate than a non-internationalized product. The isolation of message text in resource files eliminates guesswork on the part of the translator. It also eliminates re-formatting efforts caused by the translator inadvertently deleting any of the formatting codes and/or even worse, re-compilation time caused by deleting or modifying original code. The stream-lined writing style minimizes ambiguities and decreases the need for interpretation or explanations.

5 Conclusions

Internationalization is an enhancement of the production process. It includes an audit to evaluate your company's readiness and capabilities to internationalize its products, a project plan that allocates resources and establishes a process, process management that ensures communication and collaboration between all participants and the adherence to quality guidelines. The implementation of these internationalization guidelines is a first step towards a successful entry into the global market.

References

[ADAPSO 91] Preparing Software for Export. ADAPSO. 1991 (8).

[Gdaniec 93] Gdaniec, Claudia. Input-Text Identification with an Eye to MT: Translatability Index. In *Proceedings of the AMTA Workshop "Text and MT",* 1993.

[Hall 93] Hall, William S. Internationalizing Windows Software. In *Proceedings of Microsoft Windows 3.1 Developer's Workshop.* Redmond: Microsoft Press, 1993.

[Hall 94] Hall, Bill. Exploring the Windows INTL Control Panel. In *Multilingual Computing* January/February 1994.

[Horton 93] Horton, William. The Almost Universal Language: Graphics for International Documents. In Technical Communication, *Journal of the Society for Technical Communication* 40(4), 1993.

[Kumhyr & Merrill 93] Kumhyr, David and Carla Merrill. *Internationalization Report,* 1993 (Section 5.1).

[O'Hara-Devereaux & Johansen 94] O'Hara-Devereaux, Mary and Robert Johansen. *Globalwork.* San Franciso: Jossey-Bass Publishers, 1994.

[Taylor 92] Taylor, D. *Global Software.* New York: Springer-Verlag, 1992 (Chapter 2).