



Deseret Language and Linguistic Society Symposium

Volume 26 | Issue 1

Article 21

4-7-2000

Globalization and the Need for International Language Standards

Alan K. Melby

Follow this and additional works at: <https://scholarsarchive.byu.edu/dlls>

BYU ScholarsArchive Citation

Melby, Alan K. (2000) "Globalization and the Need for International Language Standards," *Deseret Language and Linguistic Society Symposium*: Vol. 26 : Iss. 1 , Article 21.

Available at: <https://scholarsarchive.byu.edu/dlls/vol26/iss1/21>

This Article is brought to you for free and open access by the Journals at BYU ScholarsArchive. It has been accepted for inclusion in Deseret Language and Linguistic Society Symposium by an authorized editor of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

Globalization and the Need for International Language Standards

Alan K. Melby

Before discussing international standards, we will explore five interrelated terms that are often used in discussions of global marketing of high-tech products: globalization, internationalization, locale, localization, and translation.

GLOBALIZATION

Globalization is the business process of making a product or service available for sale in other countries. In this paper, the focus will be on high-tech products like telephones, automobiles, DVD players, airplanes, photocopiers, and, of course, general-purpose computers and software. High-tech products typically have documentation associated with them, and this documentation can take such forms as a user's manual—either on paper or as on-line help—a technician's maintenance and repair manual, or a set of menus and messages that form the user interface for a piece of software or for a product with an embedded computer. The information in these various forms of documentation is often called "content" when distinguished from its presentation. The same content can be presented as a Word document on paper or as a Web page. The particular focus of this paper is not on products themselves, but rather on the content of their associated documentation and user interface. Thus, the standards relating to this content can be called "content globalization standards."

The trend in the world economy is clearly toward ever-increasing globalization of products. This naturally leads to the following question: When a product is marketed in a country other than its country of origin, does the product need to be modified in some way? An electrical appliance built for use in the United States cannot simply be shipped to England or France and plugged in. First of all, the plug won't fit, and secondly, the voltage will be twice that of a U.S. outlet. Unless this difference in voltage is taken into account during the product design stage, it may be dangerous to both the consumer and the product to use it in another country.

In the early 1980s, just after the IBM PC came out, I took my PC and monitor as checked luggage to London for a conference so that I could demonstrate some software I had developed. I knew I needed to deal with the voltage difference, so I took my equipment to a London computer store to get a voltage-reducing power converter. They brought out the wrong kind, and when we turned on my PC, sparks flew and my computer practically exploded. Now I can take my laptop computer to most countries, snap on one of the little adapters from my kit, and just plug it in. The computer automatically adapts to any voltage from 100 to 230 and to either of the commonly used frequencies, fifty or sixty cycles per second.

The engineering process of making a product more easily adaptable for use in other countries is called *internationalization*. My laptop required some slick engineering

to internationalize it. A low-tech product like a bag of wheat may require only a change of labels on the bag to market it in another country, while a high-tech product like a mobile phone may require considerable internationalization. Most mobile phones made for the U.S. market will not work in Europe, since they use different frequencies and codes to communicate with cellular stations. This is a source of irritation for people like me who often have business in Europe—and it gets worse. If my mobile phone would work in Spain, I would be able to take it there and use it. But a phone that has English messages that appear on its little screen would not do very well if someone tried to market it to monolingual Spaniards. And to complicate things even more, suppose you want to market the same phone in Switzerland where there are three major languages used (German, French, and Italian) and a minor fourth language. Each combination of a country and a language is called a *locale*. Thus Switzerland has three major locales: German-speaking Switzerland, French-speaking Switzerland, and Italian-speaking Switzerland. And British English is distinguished from American English as the U.K.-English locale vs. the U.S.A.-English locale.

Globalization does not, of course, imply that every product will be marketed in every locale. Part of the business process is to decide which locales to enter, and when and what aspects of a product to adjust to the peculiarities of a particular locale. The process of full adjustment is called *localization*, which includes *translation* of the documentation and adaptation, as needed, of the non-textual aspects of the product, such as its colors, symbols, and voltage. Thus, localization is a complex process that includes translation and is facilitated by prior internationalization of the product during the design phase.

Obviously, localization of a sophisticated product with considerable

documentation can be difficult and expensive. So careful thought must be put into the first issue in the globalization process: to localize or not to localize, that is the question. An interesting aspect of world languages at the beginning of the twenty-first century is that even though the use of English worldwide has increased dramatically since the end of World War II, that increase has been in the percentage of the world population that speaks English as a *second* language rather than as the *mother* tongue. In how many countries have parents switched to speaking English to their infants?

Another aspect of language attitudes is that as more nonnative speakers of English become *capable* of using products whose documentation has not been localized, fewer are *willing* to purchase products that have not been localized. This is a fascinating sociolinguistic phenomenon. In Europe, countries have gradually become more united as the Common Market has become the European Community and now the European Union. European countries have been willing to give up their border patrols and have let some of their laws become European laws. They are even giving up their national currencies in favor of the Euro. But not one country has considered giving up its native language!

The net result of these dramatic world changes is that a whole new industry, the localization industry, has appeared, mainly over the past twenty or thirty years. The need for localization is enormous. With just over six billion people on earth now (*Time Almanac* 2000, 153) and with only about 300 million native speakers of English (www.sil.org/ethnologue/top100.html) and about 300 million more nonnative speakers of English (*English Today*, 1980, extrapolated), the potential market for localized products originally in English is easily over one billion people.

The localization industry consists of three main groups: high-tech product vendors, localization service providers, and localization tool developers. The localization industry is part of what has come to be known as "the language industries." The language industries include (a) technical (as opposed to literary) writing, (b) document management (including word processing, indexing, and retrieval), (c) terminology management, (d) localization (including translation), and (e) publishing (including traditional paper publishing as well as Web publishing).

These five aspects of language industry are intertwined. A discussion of terminology management will illustrate this. One might begin by asking why terms should be managed at all. Isn't variety the spice of life, including literature? It turns out that this dictum applies only to general vocabulary, not to domain-specific terminology. In general vocabulary, you might "help a friend out of a jam" in one sentence and "extricate her from difficulty" in another to avoid saying again that you got her out of trouble. However, suppose that the maintenance manual of an automobile sometimes referred to the steering wheel as "the wheel-angle manipulator" and other times as "the manually-activated directionalizer," just to avoid continually calling it "the steering wheel." And further suppose that the "manually-activated directionalizer" also referred to the knob that adjusts the radio antenna. Variety and ambiguity can be good and interesting in general vocabulary, but in terminology, unambiguous consistency is highly prized.

New terms are being developed every day for high-tech products, and they need to be used consistently throughout the life of a document, beginning with its birth at authoring time, continuing with its venturing off into the world at publication time and its rebirth

at translation time. Today, few translations are done using typewriters. The results of *publishing* are fed back into some *document management* system for use in the localization step, and the same *terminology* used at *authoring* time must be managed in multiple languages for consistency.

THE NEED FOR LANGUAGE INDUSTRY STANDARDS

The language industries have become heavily dependent on productivity tools. Today, an accountant could not compete without the electronic spreadsheet as a productivity tool. More and more architects are using computer-aided design software as well. Likewise, the language industries depend on software tools to support their activities. Few authors would think of giving up word processing and going back to a manual typewriter. More documents are being printed from desktop publishing masters. Translators are beginning to rely on terminology management software and *translation memory* software. Translation memory is not commonly used outside the localization and translation sectors of the language industries, so further explanation is necessary.

Some localization tools are specifically designed for use with the messages and menus in computer software; others are designed for manuals or on-line help. One type of tool, sometimes confused with terminology management, is translation memory lookup. Translation memory lookup deals with segments. Typically, segments consisting of previously translated text are retrieved, along with their translation, and inserted into a document. Terminology management deals with smaller units of language, namely, individual terms that consist of one or more words that designate a single concept. Admittedly, there can be a degree of overlap between translation memory databases and terminology

databases when there are very short segments and very long terms. Terminological databases, however, typically have a much more complex internal structure and include extra information such as parts of speech, grammatical gender, indication of domain of knowledge, and definitions—none of which would be found in a typical translation memory database.

Machine translation is yet another tool used in the language industries. No single tool applies to all tasks, and that is certainly the case with machine translation, which is the process of taking a text apart, sentence by sentence, producing some representation of its structure and meaning, and formally creating a text in another language that approximates a translation of the source text.

Now, having introduced high-tech product globalization and its connection to the language industries, and having introduced some of the tools to process language, we can discuss the need for international standards relating to content globalization.

Translation memory and machine translation are often used in conjunction with terminology management. A given term in a specific domain for a specific client should be translated consistently, whether it is found in a segment of translation memory, a machine translation lexicon, or an entry in a terminology database. Spell checking in the final desktop publishing phase needs to be coordinated with the terms used in the translation phase. If all the tools used in the life of a document—including authoring, translation, and publishing tools—were part of an integrated suite developed by one company, perhaps there would be some mechanism for maintaining consistency. The fact is, however, that the state of the language industries, now and in the foreseeable future, is multiple competing tool developers. This leads directly to the need for the first of two types of standards: data exchange standards.

There are three complex types of data that flow between the tools used in the production of multilingual documents. This production flow is sometimes called the document production chain. The various tools used at various points along the chain can be called components. The three types of data are markup text, translation memory, and terminological data. We have already discussed translation memory and terminological data, but we have not yet mentioned markup. The term “markup” simply indicates the information that goes along with a text to tell how it should be presented on a screen or a page. For example, if I want to make the words “dog” and “cat” appear in bold, I might mark it up as follows:

The difference between a <bold> dog </bold> and a <bold> cat </bold> is that a dog is loyal to its master while a cat elicits its loyalty in its human house guest.

The problem is, there are many different kinds of markup, just as there are many different kinds of translation memory databases and terminology databases. The solution is to define a standard intermediate format for representing these kinds of data so that they can be passed or exchanged from one component to another without undue loss of information. A rough analogy for data exchange standards is the standards that define how a VCR, television, video game, and hi-fi set all connect together so that audio and video data can flow between components. This normally works well because of industry standards, but sometimes it breaks down. For example, if you bring back a videocassette from Europe, it will probably not play on your VCR, since Europe uses a different format for video data.

For over twenty years, I have been working with various groups to achieve widespread agreement on data exchange standards for the language industries. In the past few years, I have also been working with others on business practice standards for the translation industry, in

hopes of reducing the number of misunderstandings between requesters and suppliers of translation. In addition, I am associated with efforts to develop content markup standards. These efforts concerning the second sort of standards take patience and endurance, but some of them are paying off. A standard for exchanging translation memory data, that I helped define, has become widely accepted in the language industry. And a project called SALT, which has a mission to refine, test, and promote data exchange standards for terminological data, began in January 2000 with substantial funding from the European Union. It has subsequently obtained additional funding from the Microsoft Corporation.

The SALT project builds on the results of other projects, such as the Otelo project, which worked on the problem of allowing several machine translation systems to share the same terminology. The SALT project also builds on data exchange standards being developed within various standards organizations. I am currently involved in three of these: ISO Technical Committee 37, the OSCAR group of the Localization Industry Standards Association, and the OLIF Consortium. For more information on the SALT project, see the www.ttt.org website.

IMPLICATIONS OF CONTENT GLOBALIZATION STANDARDS

What difference does it make whether or not there are content globalization standards for the language industries? Only a few people are ever directly involved in the definition of standards. If you are not one of them, then what impact might they have on you? If you are a developer of language technology software, they will impact you in that you will have to implement them. If you work in the language industries, they will facilitate your job by making it easier

to assemble an efficient multilingual information management system in order to remain competitive. If you are an end user of products from another part of the globe, perhaps the implementation of content globalization standards will improve the quality of the content. For example, the use of terminology throughout the documentation and user interface might be more consistent. Maybe you will finally even be able to understand the instructions for programming your VCR. On second thought, that might be too much to ask. Some things in life must remain a mystery.

REFERENCES

- English today: The international review of the English language*. 1980.
- Time almanac*. 2000. New York: Houghton Mifflin.