



Home | News | Products | Services | Solutions | About IBM

ShopIBM Support Download

Search Go

IBM : [developerWorks](#) : [Web architecture overview](#) : [Library - papers](#)

Building worldwide Web sites

Make sales (or just friends) with a Web site that speaks the visitor's language

Michael Lerner
Author and President of Learn the Net.com
September 1999

*Playing in the global economy means preparing Web sites that speak to customers and communities all over the world. Find out ways to convert your Web sites to reach other cultures effectively, starting with language translation. But don't just translate -- **localize** your site. With ten tips for localizing without tears.*

English is the official language of only seven countries (Australia, Canada, Ireland, New Zealand, South Africa, United Kingdom, and the United States), comprising little more than five percent of the world's population. Yet the majority of online content is in English. While many business people speak English as a second language, most prefer to communicate in their native tongue.

As business continues to globalize, the need to localize Web sites becomes more compelling. By localizing a site into six languages other than English (Japanese, German, French, Spanish, Portuguese, and Swedish), a site can reach 90% of the online population today.

Of course the target languages depend on where a company sees its market opportunities. And it may not be necessary to localize an entire site. You can start with key pages as a baseline test to see if there's a viable market. Regardless of the route you take, new tools and services now make this challenging task more manageable.

Translation versus localization

"Translation alone does not make your site local," says Arthine van Duyne, senior producer, international, with Yahoo! Skillful *localization* -- making a site appear that it was originally developed in the target language -- depends on a keen understanding of your intended audience. Before you begin the process, consider who you want to reach. "You have to think about the user from a cultural and marketing standpoint to understand what the hot buttons are," says Ted Xistris, creative director of YAR Communications, a New York-based advertising and communications firm that has localized content for dozens of clients, including AT&T, Nike, and EDS.

For example, Americans tend to be very direct and familiar. That style and tone, however, may not work in other countries. Business conducted in Spanish-speaking countries, for instance, tends to be quite formal.

Unintended consequences

Literal translations often have unintended meaning. Brian Crouch, a business development representative with the UK-based RWS Group (see the [Resources](#) section at the end of this article) cites the Got Milk campaign. "When you translate the slogan into Spanish it means 'Are You Lactating?'" He says that it's not the job of the translator, but of the editor, to pick up on linguistic anomalies and suggest alternatives.

Humor that relies on pop cultural allusions may not "read" either, so you may have to rewrite English copy before translating it. YAR Communications, which Xistris likens to a "United Nations of advertising," has its multinational staff review materials for their appropriateness. If you work with a culturally diverse staff, you may have resources to consult in your own back yard.

Yahoo!, which has products in 19 territories and 12 languages, employs localization teams based in the target countries. The teams include producers, editors, and engineers. Van Duyne says that this is the only way to get it right. She cites one of Yahoo!'s online products -- an address book -- that sorts names alphabetically in English. In Asian countries, however, the names must be sorted according to the number of keystrokes, a discovery made by the localization team.

Human translators

Traditionally, translation has been done by trained professionals, ideally by native speakers of the target language, that is, the language the text will be translated into. That process produces the best results, but it can be time consuming and expensive. Translation service bureaus typically charge by the word. According to Karisa Stickler, a sales representative for Berlitz, the company uses a team of three people: a translator, an editor, and a proofreader. Translation into European and Asian languages of a 250-word document, about a page of text, runs around US\$250. Turnaround time is three to four business days.

Service bureau prices vary considerably, so it pays to shop around. Web-based eTranslate charges US\$50 per page for European languages and US\$60 for Asian languages. The company guarantees 72-hour delivery. eTranslate claims to have a global network of 6,000 bilingual native speakers in 80 countries who receive documents via e-mail and then perform the translations.

Uniscape says its translation management system cuts the cost of translation by capturing and reusing words and phrases. The service is something of a polyglot marketplace, allowing clients to post jobs and translators to bid on the project. By then storing the translations in a database, elements can be reused in subsequent translations, saving time and money. The company claims that its system will reduce translation costs by at least 20%.

Machine translation

Twentieth-century computer scientists have tried for decades to perfect machine translation software, with varying degrees of success. So far, convincingly accurate translation by computer remains science fiction, however. Today's machine translation systems accept text in one

Contents:

[Translate vs. localize Consequences](#)
[Human translators](#)
[Machine translation](#)
[Cultural concerns](#)
[Technical issues](#)
[Language display](#)
[Dynamic sites](#)
[Multilingual maintenance](#)
[Resources](#)
[About the author](#)

language and output in another. Rather than translating word by word, the software analyzes the text and then uses a set of grammatical rules and a database (basically a dictionary) to translate the text into the target language. The dictionary determines what terminology to use; differences in the dictionaries account for the differences between machine-translation programs. While the technology has certainly improved, accuracy varies considerably, depending on the text being translated.

Ghassan Haddad is a computational linguist who spent many years developing machine translation software. He now heads operations for Berlitz GlobalNet. Haddad says that machine translation works best for technical documents with text that follows standard English rules, with no ambiguities. "But there is no way to get machine translation to output publishable text," says Haddad. "You must have a person review and edit the text." Haddad adds, "At Berlitz we translate millions of words each year. If machine translation could improve our efficiency by even a few percent, we would use it. But we don't. That should tell you something."

Machine translations do allow readers to understand the gist of a document. If you're a developer or a researcher who needs to get an idea of what a document contains, machine translation may fill the bill. It can be also enormously handy for deciphering e-mail in foreign languages. But beware of posting machine-translated text on your site. You risk alienating your audience by providing what is perceived as a sloppy -- and possibly laughable -- translation. Still, machine translation may prove useful for some projects.

For those developers with a limited budget, Systran and Transparent Language provide a free Web-based translation service (see [Resources](#)). You may have already used Systran's Babel Fish technology for on-the-fly translation for AltaVista content. By entering text or a URL in the online text area, a translation is provided in a number of European languages. Declarative sentences translate best; translation of more complex sentences can most charitably be described as poetry. Try it yourself to determine whether it will work for you. Systran also sells translation software ranging from US\$69 for consumer to US\$975 for enterprise software.

Transparent Language, Inc. provides its free service at [FreeTranslation.com](#). The company also sells shrink-wrapped software.

To demonstrate the vagaries of machine translation, I used both Systran and FreeTranslation services to translate this sentence into Spanish: *English is the official language of only seven countries*. Here are the results:

- FreeTranslation: El inglés es el idioma oficial de sólo siete países.
- Systran: El inglés es el lenguaje oficial de solamente siete países.

While the translation of one sentence can hardly provide an accurate assessment of either program, it does demonstrate that, even with simple sentences, there are variations, primarily based on the dictionaries used by the respective products. However, a Spanish translator with Berlitz points out that the use of the word *lenguaje* in the Systran translation is incorrect as it refers to language in general, not to a specific language.

Cultural concerns

Apart from translating text, examine the site for its appropriateness for your target audience. This includes design elements, icons, navigation system, imagery, and colors. Ted Xistris of YAR Communications cites a recent project for Apple Computers promoting the iMac. He says it was not enough to just translate the product specifications; YAR had to determine which color iMac to feature in each market. In most cases plan to look to consultants in the target countries to provide guidance. "Explore each culture to see what's most impactful," advises Xistris.

Make sure also that your links speak the same language as your Web site. For instance, if your site links to English language resources and you are localizing into Japanese, you will want to find comparable Japanese Web sites or at least indicate that the site you link to is in English.

Technical issues and Unicode

Beyond the localization of the content, you will have to tackle a number of technical issues. One of the most daunting involves the use of different character sets. Western languages, such as English, French and German, use fewer than 256 characters, so they can be represented by single-byte codes. Asian languages, including Japanese and Chinese, have thousands of characters and require double-byte encoding.

Unicode, developed by the Unicode Consortium, is an emerging standard for the exchange and display of all the world's languages by computers. (See [Resources](#).) According to the organization's mission statement, "Unicode provides a consistent way of encoding multilingual plain text and brings order to a chaotic state of affairs that has made it difficult to exchange text files internationally." While ASCII uses 8 bits to represent a character, Unicode uses 16, which means it can represent as many as 65,000 characters. Unicode assigns a unique number to each character of a language. While Unicode does not currently support every language, it does cover the world's principle written languages.

The good news is that both Internet Explorer and Navigator 4.0 and later releases support Unicode encoded text and will display it properly. The bad news is that that Unicode is not a universal standard. Double-byte text encoded using other schemes requires plug-ins to display text correctly.

Right-to-left and vertical-reading language display

Further complicating the issue are languages like Arabic and Hebrew, which read from right to left. Other than Unicode, no consistent way exists for browsers to handle text display in these languages. While workarounds are available, they are platform and browser specific.

The World Wide Web Consortium is now developing standards for HTML 4.0 "to ensure that W3C's formats and protocols should be usable worldwide in all languages and in all writing systems," according to a recent Working Group statement. W3C and the Unicode Technical Committee are working closely to implement these standards.

Dynamic sites and localization

Dynamically driven sites present another problem. "The database must be compliant with binary information," says Kit Dang, an Internet Developer with Berlitz. Dang advises building compatibility into the database from the beginning, rather than having to convert it once you decide to localize the site. Other backend software may have to be adapted as well, particularly for e-commerce sites.

Finally, the home page of each language should be optimized for the search engines and directories for that particular country. You have to

determine what the most popular ones are and how they work and then adjust the metatags and keywords. "It's a time-consuming process," says creative director Ted Xistris, "but it pays off in hits to the site."

Multilingual maintenance

Once your site has been localized into a number of languages, how do you maintain them? Ideally, when you update content on the English site, you want the content on the other sites to be updated as well.

One solution is provided by Uniscape. (See [Resources](#).) Its translation management system tracks changes made to your original HTML files and then sends just the changes to the designated translators.

Enterprise software from Global Sight provides central management of multilingual Web sites. (See [Resources](#).) The software extracts the HTML code from the content. Once translations are completed, it automatically generates the HTML pages in the target languages. The software enables version control, automated workflow, and content editing. It is Unicode-enabled and works with double-byte characters. Global Sight Ambassador software starts at US\$100,000.

Brian Crouch of the RKS Group claims that it's impossible to calculate a return on investment for localizing a Web site before you begin the project. But he does say that the effort usually produces payoffs, from increasing customer service to making e-commerce sites more user-friendly. "It may take a while for the hard work to bear fruit, but it will."

Ten ways to avoid localization pitfalls

Plan ahead. Even if you won't localize your site immediately, build localization into your design specs. For instance, it's less expensive to translate HTML text than to change graphics. Saving graphics as layered files with the text on a separate layer makes them easier to alter, as does placing text *under* graphics rather than on the graphic.

All languages are not created equal. A sentence in English may be 20% longer in German and 80% longer in Hindi. Chinese is read top to bottom; Hebrew reads from right to left. Think how this might affect page layout. Flexible designs work best.

Be aware of cultural differences. Europeans typically wear white to weddings; Chinese wear white to funerals. An icon of an octagonal U.S. stop sign has no meaning in many parts of the world.

Present a clear choice. On the Web, communities are identified by language, not countries, unless the site is truly localized for a particular country. If it's not, using a flag to indicate a language can be confusing. Imagine if the Irish flag were used to represent English. It's better to list the language by name.

Know your audience. As English-speakers know, there are significant differences between American English and British English. The same holds true in other languages.

Don't sling slang. Every language uses colloquial expressions, but they don't translate easily. Of course, if your target audience expects slang, budget the funds and time to convert the text to the local argot.

Batten down the bandwidth. Many Internet users outside the U.S. have slow connections and may pay for access by the minute. So sluggish, graphics-heavy sites don't play well.

Bolster the back end. If your site uses forms, can data be input and processed in languages other than English? This may present a particular challenge with languages that use double-byte characters.

Try it out. Be sure you test the localized site with your target audience before you launch. If not, the results could be disastrous.

Keep it current. A localized site shouldn't be the forgotten sibling of the English site. Be sure to maintain it technically and update the content regularly.

Resources

- Visit the [Learn the Net.com](#), a great example of a localized Web site. The author of this article is president of Learn the Net.com, a Web-based technology training service, published in 5 languages.
- Study the [World Wide Web Consortium's emerging standards](#) for global language display.
- Find out about [Unicode](#) directly from the source.
- Try out machine translation on the Web at:
 - [Systran](#)
 - [Transparent Language, Inc's FreeTranslation.com](#)
- Put your translation project out to bid at [Uniscape](#).
- Find out more about [Global Sight multilingual management software](#).
- Find a translator through [eTranslate](#).
- Investigate the [Berlitz translation services](#).
- Check out the UK-based [RWS Group](#).

About the author

Michael Lerner is president of Learn the Net.com, a Web-based technology training service. The site is published in English, Spanish, French, German, and Italian. In addition, he writes on technology subjects for Forbes.com. He can be reached at info@learnthenet.com.

What do you think of this article?

- Killer! Good stuff So-so; not bad Needs work Lame!

Comments?

Submit feedback