The Guide to Translation and Localization

Communicating with the Global Marketplace

MultiLingual Computing, Inc.

Language Line services

LingoSystems®

POWERED BY LANGUAGE LINE SERVICES

The Guide to Translation and Localization

Communicating with the Global Marketplace

7th Edition

by

Language Line
services

LingoSystems®

POWERED BY LANGUAGE LINE SERVICES

MultiLingual
Computing, Inc.

Table of Contents

Foreword

You've heard the news:

- The number of Internet users in Europe surpasses North America by 140 million.
- Of the US population, 16% do not speak English at home.
- There are now more people online in China than in the United States.

The list goes on and on, but you get the idea. What do these facts mean to you if you want to reach a growing market? Well, that's the good news: They mean that you have started down the right path—you have picked up *The Guide to Translation and Localization*.

Once again, MultiLingual Computing, Inc. is happy to join Lingo Systems in publishing this guide. If you are new to the whole translation and localization arena, this book gives you a head start by examining industry-specific terms, choices you can make, tips on project planning, project management, quality assurance and production. It makes the whole process a little more doable without minimizing the challenges ahead. If you are an old hand in this industry, you will benefit by reviewing core concepts and reading case studies. And everyone enjoys the "Oops" entries!

At MultiLingual, we do our best to help people just like you learn how to enter the world of localization and translation, how to make the best of services and tools offered, and how to emerge with a successful launch of your international product or service. *MultiLingual*, our flagship magazine, provides a healthy mix of introductory articles for beginners, linguistic articles for translators and technical articles for engineers. If you aren't familiar with it, please take advantage of the free trial subscription offer in this guide. Our free biweekly newsletter, *MultiLingual News*, keeps people up to date with new products, services and changes in the industry. Our website (www.multilingual.com) offers a searchable database of more than 1,700 industry resources, downloadable guides, career opportunities, listings of events (such as Localization World), a blog and more.

This whole process can be quite intimidating. It should be! But the payoffs are huge:

- Research has shown that people are much more likely to buy a product if they read about it in their own language.
- Marketing to various cultural groups within the United States is a great way to expand your product reach into new areas without even crossing the border.
- The online market in China has the biggest growth potential of any online segment.

So, start reading this guide now. Check out the free information and subscription from *MultiLingual*. And good luck!

Donna Parrish
Publisher, *MultiLingual*

Donna Parrish
Publisher,
MultiLingual

President's Letter

Louis Provenzano
President and COO,
Language Line Services

For more than 25 years, Language Line Services has offered state-of-the-art language solutions to companies and government agencies across the globe, and now we are proud to be part of this state-of-the-art language resource - the seventh edition of The Guide to Translation and Localization. As the recognized standard in the translation and localization arena, the guide provides more than just information about the industry. It offers insight, perspective and critical ideas that allow companies to harness the power of language to move their products and services into new communities, cultures and countries.

With this issue, Lingo Systems and Multilingual Computing, Inc. have once again joined forces to produce The Guide to Translation and Localization, leveraging their considerable expertise in translation and language technology to reach thousands of leading companies. With the merger of Language Line Services and Lingo Systems, we are able to bring our experience to the table as well, helping to make an already powerful product even stronger. As you will see, the guide reflects the depth of technical knowledge of Multilingual's top-tier production team, the skill and diversity of Lingo System's leading translators, and the expertise of Language Line's trained interpreters.

Today, more than 176 different languages and dialects are read and spoken across the country. Cultures and languages are integrating at an increasingly rapid rate, and at the same time more products are being shipped around the globe and more services are being delivered worldwide than ever before. The importance of localizing your business, and the value of communicating in the right language, have never been greater. The guide recognizes this fact, and delivers tips for planning and executing localization projects, and case studies designed to reveal key principles and strategies for success.

Language Line Services created the over-the-phone interpretation (OPI) industry when a forward-thinking California police office saw a communication dilemma and developed a plan – and a company - to overcome it. Now, more than two decades later, the quantity and breadth of language service programs in the U.S. and abroad have exploded, with new ways of communicating coming on-line every day. The Guide to Translation and Localization makes sense of this complex new world, giving you a valuable tool for future outreach.

Sincerely,

Louis F. Provenzano
President and COO
Language Line Services

Welcome

The entire Lingo Systems team is very excited to present the seventh edition of *The Guide to Translation and Localization*. Over the last ten years we have distributed more than 25,000 copies of this book worldwide. Previous editions won several national awards from the Society for Technical Communication (STC), and we are often told by readers that they consider it to be the definitive source for information about translation and localization. While we welcome the accolades, our objectives in creating this guide remain unchanged: present clear, effective solutions that reduce costs, save time, and educate the marketplace about the localization process itself.

We are again honored to co-publish *The Guide to Translation and Localization* with **Multilingual Computing, Inc**. For those of us who dedicate our professional lives to the localization industry, Multilingual Computing, Inc. is *the* information source for the localization, internationalization, translation, and language technology industry. Their publications come in several forms, including the print and digital magazine *MultiLingual* and the biweekly e-newsletter *MultiLingual News.* They also co-sponsor the Localization World conferences. For more information, visit **www.multilingual.com**.

Since the last edition was published in 2006, the language services industry has evolved in dramatic fashion as the largest firms continue to absorb smaller ones. Lingo Systems itself has had a front row seat to this consolidation. In early 2008, we merged with Language Line Services to become the localization division of one of the world's largest language service providers (LSPs). Unlike other consolidations, however, this was not a "same silo" merger between competitors. Instead, the world-class interpretation and translation services offered by Language Line and Lingo Systems, respectively, are complementary. Language Line created the over-the-phone interpretation (OPI) industry more than 25 years ago, and today more companies trust Language Line for their language services needs (translation, localization, onsite interpretation, and OPI) than any other provider. Most importantly, together we are redefining the language service industry by offering what no other LSP can: an integrated solution to meet all of an organization's language needs, both written and spoken.

For the seventh edition of *The Guide to Translation and Localization*, we are again delighted to include chapters from some of the most respected leaders in our industry. We are confident that you will find their contributions to be both informative and entertaining. Among the distinguished experts who have contributed to this edition are:

- **Renato Beninatto**, Chief Connector for *Common Sense Advisory*, is one of the best known and most respected voices in the language services industry. Since 2002, he has led the consulting practice at Common Sense Advisory, helping global companies operationalize,

Scott Ludwigsen
President,
Lingo Systems

As a CPA, numbers are in my blood … but some mean more to me than others.

For example:

- Number of times Lingo Systems has caused a client to miss a product release date: 0

- Number of times Lingo Systems has caused a training client to miss a teach date: 0

- Average rating Lingo Systems receives on customer satisfaction surveys (1:6 scale): 5.7

- Worldwide ranking for most trusted language service provider (based on number of Language Line and Lingo Systems clients): 1

- Worldwide ranking for best language service provider: 1

OK, the last one is subjective, but the numbers speak for themselves. Our world-class team is the best in the industry and I couldn't be more honored or proud to work with them.

benchmark, optimize, and innovate best practices in translation, localization, internationalization, and globalization. Along with his senior analyst, Nataly Kelly, Renato provides his unique insights in our opening chapter entitled "The Future of Language Services" where he discusses thought-leading developments such as collaboration, crowdsourcing, web-based translation, and other technologies.

- **Donna Parrish** is the Publisher of *MultiLingual,* the leading source of information in the language services industry. In this capacity, Donna is one of the most important and knowledgeable experts on just about every aspect of our industry. Her warm introduction sets the stage for why translation and localization is important to your business and how to get the information you need to successfully connect with customers anywhere in the world.

- **Katie Botkin**, Assistant Editor for *MultiLingual*, is a relative newcomer to the localization industry who shares her "First Impressions" after just a few months on the job at *MultiLingual*. Her fresh perspective and keen observations are captured in a delightful chapter that I expect all readers, both veterans and novices, will enjoy.

- **Paul Trotter**, CEO and Founder of *Author-it Software Corporation*, explains how single-source content management tools can dramatically reduce the cost and effort needed to create, manage, and publish documentation and online content in a chapter entitled, "Single-Source Content Management Systems." Paul's chapter also describes how CMS tools such as Author-it produce materials that are more localization-friendly and less costly to translate. Lingo Systems is committed to CMS technology; in fact, *The Guide to Translation and Localization* was created and published using Author-it.

- **Uwe Muegge**—who teaches Terminology Management and Computer-Assisted Translation at the Monterey Institute of International Studies, serves on the ISO Technical Committee 37 SC3 "Computer Applications in Terminology," and is also the corporate terminologist at Medtronic, a manufacturer of medical technology—contributed a wonderful chapter entitled "Dispelling the Myths of Machine Translation." In addition to describing how this controversial technology works, he also explains how machine translation improves the productivity and consistency of human translators, how it enables the translation of material that would otherwise not be translated, and why it is not just for large organizations.

As with previous editions, the rest of the book is written entirely by Lingo Systems and Language Line personnel. Twenty-six of our team members wrote sections, or provided input, comments, and technical support. Moreover, rather than being written in "one voice," each chapter reflects the personality of its author, allowing you to develop a feel for the diversity of cultures, talents, and experience here at Lingo Systems. Pictures of the contributors are scattered throughout to let you put a face to the names you will find in the contributing writers section. Clearly this is a team effort in which we all take great pride.

The Guide to Translation and Localization is designed to be an easy read that provides guidance on all aspects of the localization process. For our readers who are new to localization, we define key terms, explain the various options for translating materials, identify issues to consider when procuring localization services, describe the localization process itself, and provide many tips to control costs and improve quality. Localization veterans will find useful information on emerging trends, as well as technical discussions about software and multimedia localization.

While we invite you to read *The Guide to Translation and Localization* in its entirety, each chapter has been written to stand alone and to be read as needed. You may find, for example, that certain chapters are relevant for your marketing staff, others for your engineering staff, and yet others for your customer service or technical publications departments. Hopefully you will find all of them helpful, with useful information for everyone in your organization.

Finally, we have again provided a number of translation mistakes in our "Oops" sections. These are funny translations that have gone awry. Many of the "Oops" were provided to us by professional linguists from around the world who were responsible for correcting bad translations done by amateurs. We expect that you will find them to be humorous, but we also want to illustrate an important point: Losing market share because your product seems "foreign" or "low quality" is not a laughing matter.

We hope you enjoy *The Guide to Translation and Localization* and find it a valuable tool for communicating with your multilingual customers and employees both in the US and throughout the world.

Sincerely,

Scott Ludwigsen
President,
Lingo Systems

Chapter 1
The Future of Language Services

by Renato Beninatto and Nataly Kelly

To make predictions about the future of language services, it is necessary to have at least one finger on the pulse of the industry as it presently stands. In this chapter, we will discuss some of the promising practices and thought-leading developments that already exist in the world of translation, localization, and interpretation, as a lens through which we can look to determine what lies ahead.

Renato Beninatto
Chief Connector,
Common Sense Advisory

Collaboration

Traditionally, translation was viewed as a craft—most translators worked in isolation. Along with changes in technology, translation evolved to segment various aspects of the process in order to harness the knowledge and skills of more individuals. This enabled translators to focus on producing translated text, while specialists were left to worry about editing and proofreading, and others, such as desktop publishing specialists and localization engineers, carried out tasks within their niches.

The next stage of evolution is to expand that talent pool even further. Today, technical and business content increases logarithmically, and every company we talk to plans to add more languages and products to their global mix. Thanks to the Web, real-time collaboration is taking off, and the standard process for the industry—translate, edit, then proofread (TEP)—is in decline for projects that require large volumes of content in a very short period of time.

Nataly Kelly
Cultural Communication
Strategist,
Common Sense Advisory

Examples of collaboration abound. Recently, Facebook conducted a crowdsourcing exercise in order to translate the content and commonly used features of its website. Rather than hire expert linguists to carry out translation, the social networking giant relied on the collective wisdom of crowds. Through a unique system that allows users to vote on different translation options, the system identifies and selects the most popular translations, to ensure that the content people see is the content people actually want. Sociolinguists are likely to smile upon this development— collaborative translation gives the power to the people and makes the selection of target language equivalents a more democratic process. It also makes an in-country review step unnecessary, because the in-country individuals reviewing the translations are already part of the process.

Collaboration is not limited to written outputs. Spoken language interpreters are also benefiting from collective knowledge, especially in situations where interpreters are providing services remotely. For example, several major telephone interpreting providers now offer their interpreters the ability to connect via a web portal, in which they can have real-time conversations with peers about terminology—all while they are actually interpreting. So, if a client suddenly uses a tricky term, an interpreter can submit a query and receive possible renditions from colleagues within a matter of seconds. When chat logs are stored, this enables companies to improve their training materials and searchable online glossaries, so that the next time a tough term comes up, the interpreter can find the term in the glossary instead.

In the future, collaboration features in translation will become more commonplace, as translation companies and buyers become more comfortable with the notion of changing their tried and true processes. We predict that collaborative translation will be more popular for some types of projects than others. For example, marketing teams already conduct testing of messages and slogans in destination markets—the power of the crowd will improve this process exponentially by enabling translation to take place concurrently with trying out different messages. For some projects, such as results of clinical trials and legal documentation, collaborative methods may be slower to catch on, as TEP may work better for projects that entail significant risk and high levels of technical vocabulary.

This means that the most important role of the language service provider (LSP) will move away from workflow tasks and into building and coordinating communities of professionals. The end result: better translations, faster.

Technology

Recent developments in technology relate directly to the trends of collaboration and crowdsourcing. Tools that are free can reach more people, but in order to gain widespread use, they also need to produce accurate and reliable results. Free web-based translation has existed for quite some time, but free machine translation that is actually worth using is a recent development. For example, Google already offers machine translation through Google Translate, but is now opening up Google Translation Center, in order to avail itself of the knowledge of masses of freelancers that may be willing to use its portal for free. In return for making this tool available at no cost, Google will obtain something very valuable—training for its machine translation tool. If the program is a success, we expect Google to significantly improve the accuracy of its automated translation technology.

As accuracy improves, such innovations can be used as productivity tools. Thankfully, you don't have to be a web mammoth like Google to make important contributions to language services technology and productivity. Companies like DotSub and SpeakLike are also using collaborative technologies to make dependable language services easier to access. Telephone interpretation companies such as Language Line Services have developed interpreting business management systems that support an array of business processes and assist companies in planning for language needs.

Changes in technology are also leading to an expansion of the portfolio of services offered. The fields of language testing, subtitling and voice-over services will evolve as these services become more accessible to larger groups of people. The interpretation market will continue growing at staggering rates, fueled in great part by language policy and immigration, but also as the result of technological advances. The outsourced interpreting market was estimated to be worth US$2.5 billion in 2007, and we predict that the over-the-phone interpretation market alone will reach US$1.2 billion by 2012. It is already possible to conference in a telephone interpreter from a mobile phone, anywhere in the world, and access an interpreter for more than 200 languages in a matter of seconds. In the future, users will have the option to decide whether they want to select the video stream, audio feed, or both.

As technology enables people to reach language services through new delivery mechanisms, the line between text-to-text, text-to-speech, and speech-to-speech services will become increasingly blurred, due to the merging of speech technology and advances in translation automation. Already, systems like ProLingua and Fluential make use of unidirectional machine interpretation (MI) and interpretation memory (IM). Right now, they represent niche markets, but going forward, these technologies will become more common. Spoken language is truly the next frontier in automating language services.

Business

The expansion of globalization and environmentalism as major trends in the world economy will generate more demand for translation. From a business perspective, the challenge is to juggle the increases in demand, scarcity of resources, and faster pace of transactions without raising prices. These business requirements depend directly upon productivity improvements— and to enhance productivity, technological advances are necessary. So, business requirements and technological advances are closely intertwined. We have argued that LSPs must take the technology bull by the horns and take on a more proactive role in developing the solutions that will push productivity to the next level.

Already, in translation management systems, we are seeing an increase in the number of systems that offer business process management features. Increasingly, companies are spotting the link between language services and key business objectives. When we talked with buyers about translation quality, they voiced strong agreement about the fact that their corporate objectives guide the way they define translation quality and their expectations of vendors. Far from being relegated to commodity status, businesses are beginning to understand that language services play an integral role in their ability to reach their concrete goals and carry out their missions.

The fluctuation of global geopolitical and economic powers is also a sign of the times. In the future, companies will need to have language services at the top of their radar in order to be responsive to such changes in international markets. Not all CEOs of Global 2000 companies currently have the words "translation" and "localization" on the tips of their tongues, but in the future, those who want to succeed will at least understand that language services are an essential part of their ability to remain competitive in a dynamic global marketplace. Those who incorporate localization into all parts of their product development will achieve the greatest results and cost savings.

Industry

The industry will keep changing in terms of its susceptibility to dynamics in the global marketplace. Prompted by currency fluctuations, various global LSPs already have been forced to make adjustments to their staffing models when employees in Europe suddenly became more expensive to use than equivalent resources based in the United States. Similarly, as middle classes grow in countries with booming economies, they may no longer prove to be low-cost centers of work for translation services, and those production hubs may migrate to new places.

Populations drift toward the money, and this has a direct impact on the language services industry. As migrant populations move around the world, the demand for language services will continue to grow. As minority language groups accumulate spending power, their ability to demand services in their languages also increases. If customers do not receive the information they need in a language they can understand, they will quickly look elsewhere to find what they need. So, companies clamoring to capture a share of those wallets, especially web-based vendors, will be at a disadvantage if they fail to make their products and services available in other languages.

The industry will also be affected by language policy. As societies grapple with changes in their ethnic and linguistic composition, they will react with legislation, policy and guidelines in order to determine how branches of government, individual agencies, and even private companies will be required to respond to the increased need for language services.

Throughout history, societies have always had the ability to communicate and interact with each other. However, thanks to technology, translingual encounters are occurring on a more frequent basis. Web-based communications have made it just as easy for a person in London to talk with someone in Beijing as it is to talk with someone in Cambridge—be that via IP telephony, an online chat room, or a simple email. The barrier of distance is, for most communications, effectively erased when the right technologies are used. Now, the challenge for the language services industry is to provide those same benefits, so that communication is quick and painless across all inputs, mediums, and outputs, and the barrier of language becomes less visible.

About Common Sense Advisory

Common Sense Advisory, Inc. is a research and consulting firm committed to improving the quality of international business and the efficiency of the online and offline operations that support it. With the goal of offering best practices and valuable insight about the localization process, our research methodology uniquely focuses on what buyers and practitioners think, do, and plan.

Common Sense Advisory
100 Merrimack St.
Lowell, MA 01852
USA
Tel: 978-275-0500

www.commonsenseadvisory.com

Katie Botkin
Assistant Editor,
MultiLingual

Chapter 2
First Impressions

by Katie Botkin

I wandered into the *MultiLingual* offices a few months ago looking for a job; I have a background in linguistics and journalism and had recently moved to the area. They hired me. Amazing, isn't it, that such a job exists—that my particular penchant for dissecting and re-assembling language, and pondering how thought translates into words, can provide an opportunity to earn money, with the added ballast of meeting like-minded individuals and learning from them.

For the first few days (let's be honest) I had some trouble pinning down what localization even was. My friends were even more confused; I gave one some online references and after much reading he came back with "I'm still not sure what it is you do."

Don Osborne describes localization quite accurately and succinctly in a personal blog entry (http://donosborn.org/blog/2008/07/09/can-we-localize-entire-libraries/) as "a diverse academic and professional language-related field covering translation, technology, and adaptation to specific markets."

Here's what I've learned so far about what that means:

Translation is only the beginning

The first time I tried to translate anything, it came out more than a little awkward, partially due to my stubbornness in keeping as many cognates intact as possible. I was about 12, and felt that some meaning might be lost if I marred the deep and impossible construction of Latin too much. My translation would be as archaic-sounding and brief as Latin itself.

This is one of the many things translators must keep in mind. One needs to consider the tone, register and resonance of the source text. However, one needs to try to mimic that in translation in a way that sounds natural to the reader of the target language. In much of localization, too, this is only the beginning. Besides the register and resonance of a target language, you have the considerations of the locale in which the translation will be marketed. Thus, for any target language, you could theoretically have many translations. Cost is always a factor, though, which may mean that clients want one and only one translation into any target language. Localization, after all, is a business.

Tools aren't as scary (or as magic) as they seem

Coming into this job, I had little to no experience with computer-assisted translation tools, except some prejudice I once got from a boyfriend who tried using them to translate some cryptic French I had posted on the Internet. Once I got past all the acronyms and terms I'd never heard, it's not so hard to understand. In fact, the whole point seems to be: get the computer to (help) translate in a way that is more sophisticated than a first-year foreign-language student, which I have been and whom I have taught. This, of course, is the real trick if you're talking about machine translation—Noam Chomsky has claimed (controversially) that humans have innate language ability; even so, most of us have trouble learning second, third, or fourth languages past a certain age. In comparison with the sum total of our experiences, computers don't process all that much data, and as a direct result (or not) are far less intelligent than we are.

Our language reflexes, however we come by them, allow for smooth flow from words to meaning to words. No good translator sits down and gives word-for-word equivalents to obtuse idiomatic phrases. No good conversationalist interprets an ironic remark literally. No good kindergarten student fails to say "please" when he or she really wants something. Language is about meaning in the real world—all of the real world. Natural language in a natural context is incredibly complex. Tools are doing a good job in their place, but they're just that—tools.

Technical English is different than literary or diplomatic English

One way to make the automation process easier on everyone concerned is to control the source texts. If you can control a natural language like English and, with style guides and company-specific terminology dictionaries, give every word or phrase one equivalent in the target language, and vice versa, then your job as a translator becomes much easier. The trend of simplified global English is taking hold; some companies are writing manuals with limited words and very short sentences. Natural language almost becomes a purely data-driven language in this case, and perhaps it should be. Technical manuals are about neither social hierarchy nor poetry, though both still enter into the equation. Getting them wrong can still be shocking or nonsensical.

Collaboration allows for some amazing work

Sharing and working together are things kids are supposed to learn to do on the playground in pre-school. Unfortunately, they also learn a lot about human nature, and human nature is, perhaps above all else, self-preserving. Sharing and working together doesn't always fly in a society of strangers, particularly if you're the only one sharing.

However, within the world of collaborative translation and crowdsourcing, sharing has reached an all-time surge in popularity and is continuing to grow. If a layman translator is passionate enough about a subject, he or she will often translate a small amount of text for free. Get enough translators doing this, and you can translate large amounts of text virtually overnight. To post-edit, open the work up for review by the rest of the crowd (as Wikipedia and Asia Online are doing, for example). Obviously, this can raise some questions about quality, but when has quality not been an issue? This is where human nature comes back into the picture: offer prizes to the crowd for the best translation or the best editing catch. Or hire a team of post-editors.

This sounds something like a pipe dream on paper, but it's working—at least on websites like Wikipedia. As everybody has probably heard by now, Wikipedia, that crowdsourced wonder, is about as accurate as the Encyclopedia Britannica.

Good or bad, everyone's buzzing about social networking

Like many people of my generation, I have my own Blogspot, LinkedIn, Facebook and MySpace pages. And like many people of my generation, I use them to write about myself, post funky photos of myself and my experiences, and to make snarky comments about other people's pages. Recently, my two younger brothers and I went to a coffee shop and sat down together—and collectively surfed each other's Facebook pages.

Everyone's buzzing about this trend and its business potential, but too few are asking if it's really such a hot idea. Advertisers are ecstatic about a market where all the research has already been done for them in the form of the hundreds of thousands of hours users have spent filling out information about themselves. But are these media-drenched kids really paying close attention to that advertising (or anything at all besides a trivial and self-referential state of being)? College courses turn to YouTube and in-class web surfing to appear savvier and to take advantage of these great new learning tools. Meanwhile, the students text message each other or listen to their iPods under their hoods, understandably bored by their 40-something teacher's attempt to relate to them.

Mark Bauerlein reports in his recent book *The Dumbest Generation* that the average young person smashes 8 1/2 hours of media consumption into 6 1/2 of actual time by multitasking, and the result is a dearth of anything substantial. Web 2.0 offers great possibilities, but at what cost? Could this be the beginning of the end for literary translation or diplomacy, except for some kind of grounded, non-web-obsessed elite? "One effect of so much social networking—so many overlapping communities of interlinked individuals—is that the language of actual human interaction begins to feel degraded. What can the word 'friend' mean after Facebook, where it is really a synonym for 'coincidence'?" Verlyn Klinkenborg asks in the July 22, 2008 *New York Times* piece "A Secret Society of 30 Million."

Whether we like the trend or not, though, it does open up the realm of "personalization." Localizing for a culture is no longer enough; you have to be able to localize for microcultures and personalities. MySpace has a site for the US Latino population. Facebook users join groups they identify with, from their hatred of Thomas Kinkaid to loyalty to certain youth hostels to love of international brands. Facebook has crowdsourced the German translation of its interface. Blogspot lets users choose from many different language formats, and LinkedIn once allowed me to track down a Chinese author I spotted on the Internet, for example. This is not to say that any of these sites will have nailed the global market; even here, cultural preference and adaptability plays a large role. *The Washington Post* recently reported, for example, that Facebook and MySpace were failing in Japan, due at least in part to shaky localization.

In many ways, professional localizers may do some groundwork here, but ultimately the idea is that each user will become his or her own localizer, choosing how and what to portray, and in what language. This is where localization answers not to the paying client as much as to the audience, since the audience largely controls it. Good or bad for our industry at large? That's yet to be seen.

The need for translation/localization keeps expanding

There are more people in the world now than 10 or 20 years ago, and certainly the world at large is more globalized—hence, global social networks and the need for translation tools and crowdsourcing. Globalization and population growth will not stop expanding, though a need for minority languages may become less and less prevalent due to this very trend of globalization.

Jeff Williams
Marketing Programs Manager

I recently connected with many of my friends from college and most of us had not seen each other for more than 25 years. We had a reunion and within minutes the years had melted away. It was like we had just said "see you soon" the day before. Past feuds and loves were put aside, and we all came together and reclaimed the spirit that united us in the first place. It was a great moment and now we're all in touch regularly instead of every 25 years. Sometimes you can go home again!

Chapter 3
Learning the Lingo

There is much confusion as to how the terms "globalization," "internationalization," "localization," "translation," and "interpretation" should be used. These terms are frequently thrown about in the press and by product developers, marketing departments, product managers, and even localization vendors. Yet understanding these terms is a critical first step when considering expansion into the competitive global marketplace and working with limited- and non-English speaking clients and employees.

Another commonly used acronym in the language industry is "*GILT*" (globalization, internationalization, localization, translation). You often see this term when the industry describes itself.

You may run into people using these terms in different ways, but here is how we define them:

Globalization
The process of conceptualizing your product line for the global marketplace so that it can be sold anywhere in the world with only minor revision. Addresses the business issues associated with launching a product globally, such as integrating localization throughout a company after proper internationalization and product design. In g11n, the common abbreviation for globalization, 11 refers to the eleven letters between the g and the n.

Internationalization
The process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for redesign. In i18n, the common abbreviation for internationalization, the 18 refers to the eighteen letters between the i and the n. See the sections on Engineering and Writing for Localization in Chapters 9 and 13 for more details on how to internationalize your products.

Localization
Adapting a software, document, or website product to various markets or localities so that it seems natural to that particular region. This may require a variety of steps including translating user interface text, modifying formats for numbers and dates, and replacing culturally inappropriate graphics or system design. In l10n, the common abbreviation for localization, the 10 refers to the ten letters between the l and the n.

Translation

The process of converting a source language into a target language. An understanding of the context or meaning of the source language must be established in order to convey the same message in the target language. Translation is a crucial component of localization.

Understanding the Differences

These four terms fit together as a "bull's eye" diagram. Globalization envelops the entire concept of taking your product line global. Internationalization is performed so that the product can then be localized. Finally, translation is the "base" component of the entire process as it represents the language transformation.

To better understand the difference between these terms, imagine that you are a product manager for a new software application that manages sales contacts. Your product development team likely assembled comments from distributors throughout the world whose customers requested new features for your yet-to-be designed contact management software. Your marketing department has determined the global demand for such a product and has developed a global branding campaign. Your design team begins work on the look and feel of the software. Here is where internationalization comes into play. You and your team must consider the following:

- Color schemes and graphic selection that avoids offending potential customers,
- Dialog boxes wide enough to accommodate text expansion,
- Functionality that supports various date, time, and currency formats,
- Input and output functionality that supports the various character sets (including double-byte characters for the Asian market),
- Right-justified text fields to prevent expanded text from overlapping the graphics, and
- A readily adaptable user interface to allow British customers to read from left to right or Arabic customers to read from right to left.

Selling your contact management software to the customers in your new markets will likely require localizing the user's manual, software, help files, and user interface from English into each target language. Fortunately, proper internationalization may lower your costs. One software manufacturer found that nearly 50% of all support costs came from consumers in foreign markets who could not understand English documentation.

Kim Villemyer
Account Manager

To say I have the career of my lifetime with Lingo Systems is an understatement! I have had the unique opportunity of working with Lingo Systems as a client and being so impressed by their service I went to work for them. Growing with Lingo Systems from its start up to one of the largest language companies in the world has been extraordinary … but working with a service team and client base I am proud to call my extended family makes working here an exceptional experience!

Other Key Terms

There are several other related terms that you should also be familiar with when you are ready to reach a multilingual audience.

Interpretation

Translation of spoken words from one language to another. Simultaneous interpretation occurs when the interpreter translates the speech as it is being given. The more traditional interpretation practice is called consecutive interpretation, wherein an entire thought is expressed by the speaker, the speaker pauses, and the interpreter converts the content for the target language speakers to hear. The skill set of an interpreter is different than that of a translator. (See Chapter 21 for more information on Interpretation.)

Content Management System (CMS)

Tools that automate the process of storing, creating, maintaining, publishing, and updating content so that users can subsequently find and retrieve large amounts of data. There are many different CMSs available on the market but many of them are not true CMSs but rather file management systems. A true CMS allows authors to place their content in a centralized system that is shared across the organization. This content can then be reused amongst authors and the content can be published to many different formats including Documentation, Help, Web, etc. (See Chapter 18 for more information on CMSs.)

Cultural Assessment

Analyzing an individual's cultural preferences through comparative analyses. Allows individuals to acquire the awareness and knowledge necessary for building effective skills and behavioral adaptations for multicultural management and business.

Cultural Orientation

Developing cultural self-awareness and effective behavioral strategies to minimize the cultural gaps that occur when contrasting value orientations of different social groups.

Translation Memory

A database in which previous translations and corresponding source text are stored for future use. New source text is automatically paired with these prior translations through the use of database technology. The matching process identifies new text, fuzzy matches and 100% matches and repetitions. The linguist uses a tool like the Trados Workbench to manage this process. (See Chapter 9 for more information on TMs.)

Translation Management System (TMS)

A category of software applications and web services that incorporates traditional translation memory functionality with workflow. Some tools on the marketplace are hosted by localization vendors, thereby blending translation services with workflow; others can be implemented as stand-alone workflow with independent translation resources. (See Chapter 19 for more information on TMSs.)

Verification Testing

Confirmation of any testable requirement, including functional testing of hardware and software system components, linguistic correctness, compatibility testing of one component to another, design verification, compliance to industry and international standards, and third party interoperability. Testing is conducted on localized, native operating systems and hardware.

Single Language Provider (SLP)

A company that specializes in translating into only one language.

Multiple Language Vendor (MLV)

A company that offers translations into multiple languages.

OOPS

At a meeting of high level diplomats:

At a meeting to discuss the relations between their two countries, the US Secretary of State, Hillary Clinton, gave the Russian Foreign Minister, Sergei Lavrov, a large red button with the word RESET in English and, above it, the word PEREGRUZKA. Unfortunately, *перегрузка* does not mean reset; it means overload. The intended word was *перезагрузка*. The diplomats were left to joke about their overloaded relations, and their future workloads.

Dan Johnson
Director of Integrated
Solutions

The global communications
industry is a wonderful
place to call home. Talk
about your insatiable
supply and demand
relationship! Humans
will always need help
communicating.

Chapter 4

What are My Options?

Executive management finally made the commitment to sell your company's products to international markets and you just accepted responsibility for managing the translation & localization process. This is an exciting time for your organization, and a significant challenge for you.

Does the project represent significant complexity and risk for you and the organization? Certainly. But the potential reward balances the risk, because providing your customers with localized products is the significant differentiator your company needs to distinguish itself in many consumer and professional technology product markets. When done correctly, it will dramatically expand your markets, increasing both revenues and profits. Your company will also develop strategic and tactical international experience that will be invaluable in the future.

So how do you identify the many options available for localizing your products or materials—and how do you select the right one for your company? After reading *The Guide to Translation and Localization*, you are ready to define the scope of your project. To ensure that your expectations are met, accurate and realistic goals for timeline, cost, and quality must be made prior to project start; and you must decide which two of these variables are most important. Once you have a clear idea regarding what is needed, where you want to go, and when you need to get there, you will be in a great position to discuss project planning with localization providers. Find out what services they offer, how much time they will take, and how much the localization will cost.

During analysis and planning, keep in mind that localization is a group effort. Most projects are relatively complex affairs that require numerous specialized resources, each functioning to provide unique and closely interrelated contributions.

The following table provides a brief summary of the different options you can select to translate your materials. Although this is not an exhaustive list of all possible alternatives, it does cover most of the common and obvious ways to localize written materials into your target languages. As you will see, you have many options for translating from Albanian to Zulu ... and everywhere else in between.

Who Does the Translation?

Beginning with the end in mind: How do you want your target audience to perceive your product in the marketplace? Is so-so good enough? If you command a premium price by offering the best product, can you afford to have second-rate documentation—in any language? Is it important that instructions for setting up and/or using your product be clear and accurate? Similarly, if you value being first to market with a new product, is it worth saving a few dollars on localization, but taking an extra couple of weeks to release it?

When delivering a fully localized product, it should not be apparent to the end user that the content they are reading or the product they are holding has been translated into their language from another. The fact that your product was originally created in English (for example) and then localized into the consumer's native language should be totally undetectable. A properly localized product should have the look and feel of having been created specifically for the target market.

If you want to produce well-localized products, then a few of the options listed above can be eliminated from your consideration. In most cases, using either machine translations or non-professional resources such as bilingual family members, acquaintances or co-workers will not produce consistently accurate, stylistically natural, or professional-quality localized content.

The use of locally based, single-language translators can be an effective solution for small projects with one target language. This option works best when translating from English into a common language (so that it is easy to find a linguist), the formatting is simple (such as Microsoft Word), subsequent updates are unlikely, timelines are flexible, and projects are infrequent. As the volume of material and/or the number of target languages increases, the limitations of this approach will become increasingly obvious. Keep in mind that with single-language translators, your management and coordination load will increase exponentially as the number of target languages grows. Furthermore, the efficiency and cost savings gained through the use of a dynamic, multi-step localization process and sophisticated localization tools (discussed in Chapters 6 and 9) will become more important as well.

OOPS

From a UN interpreter:
We need to grab the bull by something other than the horns.

Localization Options	Characteristics	Benefits	Limitations	Considerations
Acquaintance or family friend	Someone you know who happens to speak or has studied the target language	Very low or no cost	▪ Not a localization professional ▪ Limited capacity ▪ No QA steps ▪ No tools ▪ No process ▪ No technical expertise ▪ Extended timelines ▪ Consistency and quality are likely to be poor ▪ High risk of missing deadlines ▪ Hidden costs ▪ Updates are time consuming	▪ Speaking a foreign language does not qualify someone as a translator ▪ You get what you pay for
Bilingual employee	Someone in your company who speaks the target language	▪ Low out-of-pocket cost ▪ May have subject matter expertise	▪ Not a localization professional ▪ Limited capacity ▪ No QA steps ▪ No tools ▪ No process ▪ Timelines uncertain ▪ Consistency and quality are likely to be poor ▪ High risk of missing deadlines ▪ Hidden costs ▪ Updates are expensive	▪ Competing responsibilities at work ▪ Will they be available when you need them? ▪ Using in-house resources to review the final deliverable is highly recommended
Overseas distributor	▪ Foreign distributor, agent or representative ▪ Translation 'cost' may be negotiated as part of overall sales agreement	▪ Low out-of-pocket cost ▪ May have subject matter expertise ▪ Someone else is responsible for project	▪ Not a localization professional ▪ Limited capacity ▪ No QA steps ▪ No tools ▪ No process ▪ Extended timelines ▪ Consistency and quality may be poor ▪ High risk of missing deadlines ▪ Hidden costs ▪ Updates are expensive	▪ Should be negotiated "up front" when the distribution agreement is prepared ▪ You may lose control over content and quality ▪ Ownership, copyright, and IP issues are all a concern ▪ Unauthorized changes might be made and go undetected ▪ Content consistency between different target languages can be difficult to maintain

Online machine translation	Just to get the 'gist' or basic idea of written text	Free, often humorous results	Machine translate the written text back into the source language for more humorous results	Proceed with caution; you get what you pay for
Crowdsourcing/ other emerging solutions	▪ Goes out to a community to be translated (usually online)	▪ Cheap, free	▪ No control over translation ▪ Few quality controls	▪ Early-stage technology ▪ Let someone else stumble and perfect before going this route
Individual translator	▪ Independent contractor ▪ Specializes in one language ▪ Is often locally available	▪ Translation professional ▪ Easy access ▪ Quick turnaround on small projects	▪ May not have translation memory and other sophisticated tools ▪ Limited capacity ▪ Longer timelines on larger projects ▪ No independent QA (if any) ▪ Quality may be an issue ▪ Updates can be expensive ▪ May have no DTP capability ▪ One language pair only	▪ For small, single language projects this might be the right solution ▪ Expect to provide a lot of project management support for ongoing single or multiple language translation efforts ▪ With only one linguistic step, quality may be substandard
Full-Service multiple-language vendor	An organization dedicated to providing a broad range of linguistic services	▪ Localization professionals ▪ Unlimited capacity ▪ High quality & consistency ▪ Aggressive deadlines can be met ▪ Multiple QA steps ▪ Sophisticated tools, TMs ▪ Technical expertise ▪ Updates are quick and cheap ▪ Proven process ▪ Any number of languages ▪ Project management provided ▪ Full range of localization services	▪ Can be overkill for projects with extremely limited scope	▪ Use the right resources for the right job ▪ Outsourcing localization allows you to focus on what you do best

Deborah King
QA Specialist

The proofreading compulsion is part of a larger syndrome with many interrelated symptoms. I am happiest when my desktop is empty and my wastepaper basket is full. I love to delete my email inbox. Nothing puts me in a better mood than correcting menus at fancy restaurants and handing them to the maitre d' on my way out the door. What can I say? There's no twelve-step program for this.

Domestic vs. Overseas

You might be tempted to use your overseas office to localize your product. They speak the language after all, so it should be easy, right? The temptation is even greater if you have an in-country subsidiary or distributor offering to do the translation for you. While it is true that these options may be the best solution in some cases, it can also lead to other problems:

- Less control from headquarters,
- Difficulty in project coordination and communication (if nothing else, the time zones will be an issue),
- Unauthorized changes to the content,
- Risks to schedule, and
- Incorrect translations (they are likely not professional translators).

When you send your materials to an overseas office or distributor, you create an opportunity for them to modify both your content and message. The in-country team may have different priorities from your US-based team, resulting in changes to branding, use of terminology, and perhaps features that have been disabled or removed from the US version.

Unfortunately, you might not become aware of these modifications until a problem arises or someone translates the in-country translations back to English. At Lingo Systems, we have a client that took this route and was alarmed to discover that their foreign office had deleted all contact information for in-country technical support and instead substituted a comment instructing the consumer to contact the reseller with any problems!

Stories such as these can be avoided by choosing the right people for the right job. In most cases, the best result is obtained by hiring a professional localization resource for translation and then using your in-country representatives for terminology list development and final review of the localized content. This strategy enables you to coordinate the localization efforts centrally, while encouraging your overseas partners to buy in on the final product.

Your final choice is between hiring individual translators or a full-service localization vendor to manage all of your localization efforts. The considerations here involve time, quality, budget, and the need for value-added services. Do you have the time and staff to hire and manage translators and assess the quality of their results? A full-service vendor can provide you with all the resources necessary to receive high-quality translations on time and on budget, reducing your need to be involved in the day-to-day execution of the project.

As you consider all of the available options, let your project goals lead you to the best solution.

Chapter 5
Planning a Successful Localization Project

You have received the green light for your product or service to go global and now you must decide how to manage the project. There are several options for localizing your materials. Selecting the correct resource, however, is only the first of many issues that will need to be addressed in order to choose the best solution. The answers to some "big picture" questions will affect your localization choices. Before you start, ask:

- What is your long-term globalization strategy?
- How is your company positioned? Is it a market leader or specialty supplier; top-of-the line standard-setter or low-cost alternative; custom manufacturer or commodity production; high-quality provider or low-quality solution?

Other basic questions that are important to answer are:

- Which products and components will you localize?
- What target markets and languages do you need?
- What are the legal, regulatory, liability, and commercial requirements in the target market(s)?
- What is your timeline?
- What is your budget?
- What level of quality and consistency will you need?
- What is the likelihood and extent of on-going future updates?
- How often will you have new products for localization?
- Are there engineering and functionality considerations?
- Are there specific requirements or guidelines that need to be incorporated into the process?
- Are the source formats compatible with the languages targeted for localization? If not, will conversion options be considered?
- Is voice-over required, and if so, what are the gender requirements and audio specifications? Will the audio be timed to a video?
- If the final product is an interactive website or application, how much functionality testing will be required? Are test scripts available, or will they need to be developed prior to testing?

With so many questions to answer, perhaps the question of paramount importance is, "How are you going to do it?" *The Guide to Translation and Localization* has this answer and many more.

Diana Joyce
Operations Coordinator

As the Operations Coordinator for Lingo Systems, I am backed by a strong team that is responsible for planning, project management, desktop publishing, and quality assurance. I have been with Lingo Systems since 2005 and prior to assuming my current position I served as an Account Manager and was very involved with all aspects of the sales process. I hail from Southern California and have raised two successful young men with my husband, Jim.

How Much to Localize?

The first question many companies ask is, "How many components should I localize?" The answer can be anything from "only the essentials" to "all content for all product components." The second question often asked is, "How much can I afford to localize?" When trying to answer the money question, however, a better question might be, "What is the financial impact of not doing it?" By choosing not to localize certain products, you run the risk of missing potential sales, or even worse, offending the target consumers in a new market by not providing information in their language. Depending on the product, you could also find yourself in violation of various legal or regulatory requirements, possibly even preventing distribution of the product in the target market. It is strongly recommended that you seek counsel regarding the legal implications of not localizing some or all content when expanding into a new market.

Unfortunately, when evaluating the business case for localization, many companies only factor in how much it will cost. Do not make this mistake too! An equally important question is, "How much additional revenue can we reasonably expect to generate?" In our experience, your international sales and marketing staff will most often drive the commercial justification to localize.

Another item to consider when deciding how much to localize is whether economies of scale are available. Once a localization program is established for one language, the file preparation, project management, and back-office administration are already established. Localizing into additional languages can take advantage of these initial investments.

Finally, depending on the nature of your products or services, you may not have much choice in whether to localize or not. For example, if software developers want to increase customer usability, in addition to localizing the software interface, they will need to consider localizing the application itself. This includes help files, "read me" files, installers, legal warranties, user guides, and installation instructions as well. Other examples include developers of training materials who must decide which courses should be taught in which languages and consumer electronics manufacturers who need to evaluate the merits of supplementing user documentation with customer support websites and other online resources.

Fortunately, localization of virtually any component is straightforward with the right planning and the right technology. Several tools and methodologies have been developed that allow you and your localization partner to "recycle" translated content across many different media types. These include translation memory databases (discussed in Chapter 9) and single source content management systems (discussed in Chapters 12, 17, and 18). By reducing the volume of new text that needs to be localized, and keeping changes to a minimum, costs are reduced, timelines are shortened, and consistency is improved across all components.

Plan Early

There is an old carpenter's saying: "Measure twice, cut once." In other words, plan carefully and early for localization. For example, if your documentation includes 20 screen captures from the software user interface (UI), the UI should be localized before the documentation so that terminology that is frozen and translated in the UI can be propagated to the documentation and/or help files.

When tight timelines require that UI and documentation localization occur simultaneously, aggressive localization schedules may still be possible. Because these projects frequently involve the translation of thousands of words, vendors form teams of linguists to work on both components at the same time. At some point in the schedule, after the UI is frozen, time is allocated to allow the documentation to "catch up" so that references to buttons, menus, and other items in the text of the documentation match the terminology used in the UI. Fortunately, other components such as training materials and web content can usually wait until the bulk of the product localization is complete. After all, end users can't be trained until you have something on which to train them!

One Language at a Time or All at Once?

Many international companies prefer to roll out new products to all of their markets in a simultaneous or "simship" release. Although such releases are a goal of some multinationals, they are not a commercial priority for everyone.

A simultaneous release poses two main challenges for localization. First, in order for a company to release source and localized products at the same time, localization generally needs to begin while the source is still under development. This means that each last minute change to the UI, online help, or other documentation must also be incorporated by the localization team. As you can imagine, such "stops and starts" make configuration management more complicated and the project more expensive.

Given these challenges, consider an iterative development life cycle where your localization vendor is provided with the "alpha" or "beta" version of the software. Later, when the product is "functionally complete," the vendor can finalize the translations. This approach means a little more work, but everything can be finished for a "simship." Alternatively, if you choose a delayed release—localizing your components as they are needed—you can lay the groundwork with your vendor so that each component is "ready to roll" through the production process on your signal.

The second main challenge involves managing localization team complexity and resource limitations. Depending on the word count, timeline, and number of languages, your localization vendor may need to assemble teams

of translators, copy editors, and proofreaders to translate the content. There may also be several teams of desktop publishers to lay out each page; two, three, four, or more QA reviewers to inspect the work as it is done; and multiple localization engineers processing and preparing the files for each step. Coordinating all of this activity requires one or more project managers. At Lingo Systems, we have delivered large rollouts to 30+ countries that required a team of more than 100 professionals! When qualifying a localization vendor, make sure they have the resources and experience to handle your project.

Getting an Estimate

You have done your homework. You have scoured the latest issues of *MultiLingual*. You have spoken to your counterparts in the Society for Technical Communication. You may have even checked your local telephone directory, or perhaps conducted a web search. And, after compiling a list of potential vendors, you want to hear what they can do for you.

As you begin your discussions, you may find that you have many more questions than answers. Undoubtedly, you will be asked to clearly articulate your requirements and provide electronic source files for the vendors to analyze. This stage is all about exchanging information. What information does your vendor need? What information do you want back? Depending on the size of your project, you may even consider following a formal RFI (request for information), RFP (request for proposal) or RFQ (request for quote) process. Remember, assumptions represent risk for both you and the vendors. So, the more information you provide and the clearer and more concise your instructions, the more accurate the estimate and more realistic the project plan will be in return.

To help differentiate between vendors, many companies ask for a sample translation as part of the bidding process. This can be a useful tool if your materials are highly technical and you want to ensure that the localization provider is qualified to handle the translations with linguists that have the appropriate subject matter expertise. A word of caution: Since localization providers want to make a good impression, they will most likely use their best linguist to translate your sample. Unfortunately, it does not necessarily follow that the same linguist will be available (or utilized) if the vendor is awarded the work.

Asking for references can be a fantastic way to evaluate and compare potential vendors. As with sample translations, you will almost assuredly be provided with contacts who will provide positive feedback, so your objective should be to get a feel for the style and strengths of each vendor in order to determine which one will be the best fit for your company. Start by asking for companies whose projects were similar in size, scope, and type

Eric Manning
Project Manager Supervisor

I've been with Lingo Systems for 8 years, and I'm still learning. As soon as you think you know everything you need to know about localization, a client sends you source files in a format you've never seen before, or some company releases yet another tool to master. I currently live in East Portland with my wife, 3 girls, 3 black cats, and a hyperactive Cavachon puppy. Sleep? There'll be plenty of time for that when I'm dead.

to yours. Then, when speaking with the references, ask them to describe their experience working with the vendor: what was helpful, how long did the project(s) take, and how was the customer service? Find out how long the references have worked with the vendor. Reliability and long-term consistency will be important factors in your selection.

Selecting a Vendor

Unless you provide a template with your instructions, you will likely find that each localization provider has a slightly different way of presenting their estimate to you. Some vendors will respond with a great deal of information detailing the specific tasks they propose to perform, the amount of effort that is required, and a business case for why they are the best choice to perform the work. Others may only submit a standard form with minimal detail.

The process that each vendor proposes to follow on your project, the number of linguistic and QA steps employed, the qualifications and locations of the resources, and the tools that they use vary significantly. These differences can have a huge impact on cost. But how do you choose between competing proposals when one vendor's estimate is 20% higher for what amounts to 50% more effort? Similarly, how much is it worth to you if one vendor uses college students to perform the translation while another only employs full-time, professional linguists?

oops

Packaging on a Turkish dessert:

"Nutrition Useful: Its rich with the protein, fat, B vitaminphosphorus and potassium. It affected as a blood maker, and nutritive forebrain cells. It reduce the cronerheart diseases and cancer risk, because containing of E vitamins, unsaturated fat and flovanoids. Its perfect source of energy." Stare in a coll dry place away from strong light & odours."

Dougal Cullen
Senior Project Manager

Managing a global team on a 24-hour clock has its advantages; when you're up all night there is always someone to talk to!

Obtain the best value for your dollar by making an *apples-to-apples* comparison, but be forewarned: This is easier said than done. Standardizing localization estimates can be difficult and time-consuming. As with most things in life, you get what you pay for, and localization is no exception. Take the time to investigate each vendor's services thoroughly. Begin by asking a potential vendor these questions:

- What subject matter and industry experience do you have?
- How do you qualify your linguists?
- Who would manage my project, where would they be located, and what would be the process?
- If required, would you be willing and do you have any experience staffing a project manager onsite at the client's location?
- How often would I receive status reports on my project and what information would be in them?
- Who would be my primary contact during the project?
- Do you use state-of-the-art localization tools and, if so, which ones?
- When analyzing my electronic source files, do you calculate leveraging?
- How do you charge for repetitions, fuzzy matches, and unique text?
- Will you create a TM (translation memory database) and will I own it?
- How many projects like mine have you managed before?
- Can I speak with your previous clients about their experiences?
- How many linguistic steps will you perform?
- What is your quality assurance process?
- How would you develop and maintain a terminology list specific to my project and/or industry?
- What is your process for incorporating changes from our in-country review? Do you charge extra for this, and if so, how are those costs determined?
- Would my in-country team be able to speak directly with your linguists if necessary?
- How would changes be handled during the course of a project?
- What is your record for delivering on time?
- Is your estimate firm or is it subject to change?

A qualified localization vendor should be able to provide an estimate that is comprehensive, accurate, and clearly defined. The table on the next page offers a sample pricing structure covering various services and how they might be billed.

Project Type	Task	How it is billed
All Projects	Project management	Typically 10-15% of the project costs
	Translation, new text	Per word (usually a minimum charge if <250 words)
	Translation, fuzzy matched text	Per word, normally less than full word rate
	Translation, 100% matched or repetitive text	Per word, normally less than full word rate
	Copyediting	Per word or per hour (usually a minimum charge if <950 words)
	Proofreading (documentation) or online review (software or web)	Per word or per hour (usually a minimum charge if <2,000 words)
	Glossary/terminology list development	Per term or per hour
	Translation memory creation, administration, and updating	Per hour
	File treatments/file prep	Per hour
Documentation	Desktop publishing	Per hour or per page
	Graphic design	Per hour or per graphic
	PDF creation (Print or Functional)	Per hour or per page
	Quality assurance (QA)	Per hour
Software, website, & online help	Desktop publishing	Same as for documentation
	Help generation & QA	Per hour
	Engineering	Per hour
	Functional testing	Per hour
	Graphics and screen captures	Per hour
Voice-over	Voice talent	Per hour (usually a minimum charge if <2 hours)
	Studio time (audio recording, editing, archiving)	Per hour (usually a minimum charge if <2 hours)
Transcription	Conversion of audio to a written script	Per minute of recorded audio. Dependent on number of speakers, language(s), audio quality and level of transcription detail required.

Kaysi Cullimore
Senior Project Manager

Witty line of text? Not from me—but I'd be happy to make a project schedule instead.

Selecting the right localization partner is a critical component to the overall success of your expansion into new international or multi-lingual markets. The value of developing a long-term partnership cannot be overemphasized. Such a relationship provides a means for the localization team to learn about your company, constituents, and products, and to understand your requirements and expectations inside and out. Often this results in improved workflows, shorter timelines, and lower costs for you. The better your partner understands you and your product line, the more smoothly the localization process can proceed, and the more effectively project management and communication protocols can be fine-tuned. The long-term relationship between you and your localization provider is, ultimately, the best way to achieve cost-effective, high-quality work for each and every project.

OOPS

On a road sign in Wales:

All official road signs in Wales are bilingual, so the local authority e-mailed its in-house translation service for the Welsh version of: "No entry for heavy goods vehicles. Residential site only." The reply duly came back and officials set the wheels in motion to create the large sign in both languages. The translation in Welsh said, "I am not in the office at the moment. Send any work to be translated."

Chapter 6
Process is the Key to Success

What constitutes a successful localization project? For the client, it probably means receiving high-quality deliverables, on time and on budget (or, even better, early and under budget!). For the localization provider, it should mean producing profitable, high-quality deliverables on time, every time. The similarity of goals for both the client and the vendor is clear, and this overlapping definition of success has created a central truth in localization: Process is the key to success.

To remain competitive in the global marketplace (or perhaps more appropriately, the online marketplace) companies are localizing more products and materials than ever. To effectively manage and ensure consistent, high-quality results, your localization vendor needs well-defined processes and workflows. Translations used to be simple: the client wrote a technical publication, sent it out to a localization vendor, and received the target language version(s) some weeks later. Today, companies of all sizes realize that they can grow, increase revenue, and maximize the return on their localization investment by communicating with their customers in their native language in as many different ways as possible, including:

- GUI,
- User manuals,
- Service manuals,
- Online help,
- Company websites (often featuring Flash intros, impact movies, and other multimedia tools),
- Virtual storefronts,
- CD versions of Help (included with the product or as part of the software),
- Packaging materials (box art, labels, inserts, envelops, etc),
- And more!

Delivering all of these components as part of a simultaneous, world-wide rollout is a very complex undertaking. Fortunately, if you select the right vendor, they will simplify your localization project for you because they will follow a proven, effective, and efficient process that yields consistent, high-quality deliverables.

Mary Pursifull
Project Manager

I was born near the village of Pepin, Wisconsin, in what was then known as the "Big Woods" of Wisconsin. There wasn't much at all to do around those parts except read, and when I picked up Noam Chomsky's book *Language and the Study of Mind* my love of words and language was set for life. This love of language makes every day at Lingo a wonderful experience.

Not All Processes are Created Equal

Vendors can vary the number of linguistic steps they perform when translating your materials. Some employ a single step, others two, and the best, three. Vendors also have a choice when selecting the linguists who will perform the translations. Some try to save costs by hiring college students or other non-professionals who happen to speak the language and want to earn a few extra bucks working part-time in the evening. Others employ on-site linguists or form long-term relationships with in-country language professionals who work full time at their chosen profession. Finally, the quality assurance reviews performed by localization providers may be non-existent, minimal, or thorough, with the best ones utilizing regression checks and other statistical procedures.

At Lingo Systems, a typical documentation project will include 20+ process steps and even simple projects have at least 14 separate steps. Because no two projects are the same, there is no one standard workflow that can be applied. Instead, we customize the actual number and type of steps during the planning phase to ensure that clients receive the best solution based on their requirements and budget.

The steps we perform fall within one of the following eight disciplines. Each is a unique specialization that is ideally performed by skilled professionals who are specifically trained in that area.

- Project Management
- Translation
- Copyediting
- Proofreading
- Engineering
- Desktop Publishing (DTP)
- Quality Assurance (QA)
- In-Country Review (ICR)

The need to effectively coordinate and execute these tasks is why process is the key to success. A good vendor will have well-documented and proven processes for all types of projects and files: documentation, software, help, animated movies with voice-overs, etc. Such processes promote cost effectiveness, consistency, quality, and flexibility.

Costs are controlled because each member of the team is clear on the tasks they need to complete and the timeframe allotted to do so. Consistency is achieved because assigned tasks are executed according to the required standards the same way each time. And perhaps most importantly, by having a documented plan there is a reliable starting point for problem-solving and adapting to a variety of project requirements.

Role of Technology

Technology requirements play an increasingly important role in localization workflows. Whether your company is in the Fortune 500 or a grassroots start-up, choose a localization vendor that uses advanced localization technology to help you cut costs, improve consistency and quality, and increase long-term value by properly maintaining your translation assets. Automated workflows, CMS integration and customization, and single-point, web-enabled access can decrease the chance for miscommunication. Additionally, these systems can drastically reduce timeframes for file transfers between parties across time zones.

As more content is translated, translation memories become larger, terminology management becomes more complex, and both form critical components in the process. Multilingual terminology lists and glossaries form the foundation of good translation memories. Together, these components are the centerpiece of a company's localized intellectual property. Ensure that your vendor has the knowledge and expertise to properly maintain and protect these valuable assets. Plan ahead and verify that the vendor you choose has the resources to meet your future needs and the expertise to use the latest technology and specialized tools to make your localization projects a success.

Communication

Good communication is critical to successful project management in any field, and it is at the core of a good localization process as well. The project manager (PM) must effectively communicate your needs to the localization team working on your project and similarly communicate issues they identify back to you and your colleagues.

Typically, employees from both the client and the localization provider work together throughout the project, with the PM serving as the main communication conduit. A primary contact should be identified on the client side to coordinate communication. Ideally, this person has the authority to approve and/or delegate tasks. An additional point of contact should be identified as an in-country reviewer who resides in the target country. This person offers advice on terminology, style and local technical issues, and performs quality reviews. On the localization provider side, the PM coordinates directly with all project resources, including linguists, software engineers, desktop publishers, and quality assurance specialists. Any issues raised by a member of the localization team are funneled back to the client by the PM for resolution.

Clearly communicate your expected delivery dates, delivery format, and media at the beginning of the project. Similarly, your localization team should confirm their understanding of your requirements to ensure that everyone is on the same page. On many projects, the localization provider

Olivia Halfen
Senior Project Manager

"De son schenkt sching" …
Say that to anyone and
it sounds like you are
speaking an Asian language.
Say that to someone from
my hometown, Utscheid,
Germany, and they will
pour you a beer.

determines the timeline at the project start based upon your specifications and the availability of linguistic and technical resources. In this situation, the PM should also confirm the delivery date with you so that your expectations are met.

Remember: You know your product better than anyone, so communicate as much information as possible to your localization vendor before your project starts. This includes anything that may be helpful for translators to better understand your product, such as existing glossaries, terminology lists, or product descriptions.

In addition, your localization provider should provide you with clearly defined status reports and/or reporting analytics. Status reports can be as informal as an email, or as formal as a posted report via a user interface on a website. Be sure to request summary reports that meet your specific needs. The largest vendors may also offer a translation management system (TMS) that allows clients to view the status of their projects 24/7. See Chapter 19 for a more in-depth discussion of TMS tools.

During the Project

The PM assigned by your localization vendor coordinates all resources, monitors progress, troubleshoots issues, and provides necessary information for all resources to successfully complete the project on time. In addition, the PM informs you of the status of your project based on your requested method and frequency. Aside from these status reports, you may not hear a lot from your localization provider. Do not be alarmed! Rest assured that the localization team is hard at work completing your project.

In Lingo Systems' view, a high-quality translation process should involve three distinct linguistic steps: translation, copyediting, and proofreading. Depending on your word count, this could be as short as a couple of days or as long as several months. Note that some providers may try to combine the translation and copyediting phases into a single step. This can reduce costs and shorten timelines but it risks reducing quality and consistency. Be sure you know how many steps your provider will perform and be certain to document it in your contract or work order.

After translation, your localization team sends your files through desktop publishing, engineering, and quality assurance stages. The number of people involved and the type of work done to your files is predicated on the type of materials being localized. A small marketing flyer translated into a single language would likely require a six-person team: three linguists, one each for translation, copyediting, and proofreading, a desktop publisher, a quality assurance specialist, and the PM.

Alternatively, a large operator's manual translated into multiple languages will require a much larger and complex team. Depending on the word count, timeline, and number of languages, your localization vendor may assemble multiple translators and editors to process the word count more quickly while still using a single proofreader, who acts as a gatekeeper for consistency and quality. In addition to the linguists, there may be several teams of desktop publishers laying out the pages of the manual; two, three, four or more QA reviewers inspecting the work as it is done; and multiple localization engineers processing and preparing all the files for each step. All of this activity will be coordinated by a key, central figure: the PM.

As you can see, depending on the type of project, there may only be a handful of people who touch your project, or a whole armada! At Lingo Systems, we have the expertise to simultaneously roll out translations in 50 or more languages requiring teams of more than 100 professionals! Regardless of the type of project or the size of the team, it is important to select a vendor with an established, effective, formalized process that is always followed to ensure good communication, quality, and consistency.

A good process can also mitigate disruption if you need to change or modify the source document after a project has already started. However, there are several important issues to keep in mind. Frequent changes during the course of a project can be expensive and can severely impact the original timeline. In the event that changes are required mid-project, clearly communicate the details of these changes so that your PM can quickly incorporate them into the final product. Changes to the original project scope may cause an increase in localization costs and/or a delay in delivery. Have your localization provider submit an estimate for each new change, subject to your approval. This estimate should address both additional costs and any delivery date adjustments that may be necessary prior to incorporating the changes.

Completing the Project

For many projects, an in-country review (ICR) step is performed before the final files are delivered. This is an opportunity for one of your colleagues or business partners in the target market to review the localized files and provide expert feedback to your vendor. An ICR accomplishes several things: prior to release, it puts the files in front of an expert with first-hand knowledge of the material; it creates an opportunity to gain buy-in from your international partners; and it can increase your confidence in the quality of your vendor's work. Read more about these in-country reviews in Chapter 7.

David Morrow
Account Manager

There's this book, *Everything I Ever Needed To Know I Learned in Kindergarten*. Not true! Growing up we are constantly learning how to effectively connect with one another. Communicating to one person or a million people anywhere on the globe or right next door, to folks who represent a variety of cultures, languages and customs—what a great industry to be a part of!

Providing feedback (both positive and negative) to the localization vendor is an important closing step to every project. Through this feedback, you help your vendor define and make process improvements that will help improve the quality of future deliverables. Offer a candid critique of areas where your vendor failed to meet your expectations and praise for tasks well done.

A lot of time, money, and resources go into a localization project, so the more you leverage the resources of a qualified vendor with an established process, the more likely it is you'll be satisfied with the outcome. By using a well-documented and proven process, your localization provider will have the framework and controls in place to ensure a positive experience for you. Process really is the key to making sure you receive the highest quality localization service.

oops

Ireland's worst driver

In Ireland, a 2 year search has been out for the nation's worst driver, a Polish driver named "Prawo Jazdy" who has acquired roughly 50 delinquent speeding tickets, each filed under a different home address for the driver. After 'careful research,' the Irish Police discovered the riddle that troubled them for so long. Prawo Jazdy is not a person; it's the Polish translation for "Driver's License." Instead of addressing tickets to the individual's name, they addressed them to "Driver's License" which can be found at the top of the Card.

Chapter 7
Producing Native Quality Content

In the global network of localization, there are thousands of vendors, agencies, and linguists all claiming that "quality is job one." No doubt you have heard that tagline before. But what is quality? Has it been reduced to little more than a tired marketing phrase? Perhaps, but there are quantifiable methods for assessing quality so that it becomes empirical and not ethereal. In this chapter we outline the steps and strategies you can, and should, employ to ensure the highest linguistic quality for your localization project.

So just what does quality mean when it comes to localization? Some people translate instructions for the proper way to install a defibrillating device in a patient's chest. Others translate materials that will be used only once during a training class. And a few just want to know what the email from their Turkish distributor says. From a practical standpoint, each of these scenarios requires a different level of quality. To determine the appropriate level for your project, first identify what your needs are, how your materials will be used, and who your audience is. Then, communicate your requirements to your vendor so that they can deliver the quality that you want and need. Once you have defined it, you will be able to ask for it, and then you will know it when you see it.

The quality of your final localized deliverable is largely dependent upon the quality of the original source text. To put it another way, "garbage in, garbage out." A poorly written source document will appear as a poorly translated target document, regardless of the quality of the translators.

Before your source materials ever go to translation, they should be "localization ready." The easiest way to ensure this is to develop them with localization in mind. If your technical writers need assistance, put them in touch with your localization vendor. Alternatively, if the materials have already been developed, your localization vendor can assist with internationalizing them. Either way, you will have source materials that are ready to be translated the right way. This is the first step toward obtaining quality translations. Other important steps involve:

- Providing reference materials, such as style guidelines, a glossary in the source language, and a terminology list in the target language,
- Selecting linguists,
- Using a three-step translation process, and
- Arranging for an in-country review.

Jennie Jorgens
Administrative and
Accounting Assistant

I have a glass desk at home
with the word LOVE
written on it in many
different languages …
love, amour, liebe, amore,
kærlighed, rakkaus, szeretet,
amor, láska, ljubezen, kärlek.
What would our world be
like if we lived as if we truly
loved one another?

Style Guidelines

Style guidelines, or style sheets, are a list of specific "rules" for the linguist to follow during the translation process. You may already have a corporate style guide for your source language, as well as for target languages you have previously localized. If not, your localization vendor can help you create them. Developing a style guide begins with a generic style guide for each language; additional corporate and project-specific rules are then added and will take precedence over the generic rules. Style guidelines typically address the following issues:

- Desired tone (formal vs. conversational) of the localized documentation,
- List of terms that should be translated, and those that should not ("Do Not Translate" list),
- Rules for capitalization, spelling of numbers, punctuation, and accent marks,
- Translation of titles and subtitles,
- Conversion of measurements,
- List of dos and don'ts,
- Use of abbreviations, and
- Cultural conventions, considerations, and taboos.

Style guidelines, developed in consensus with all stakeholders, help create high-quality documents appropriate for the end-user, for meeting company and country standards, and for maintaining geographic and cultural suitability.

Glossary

A glossary is a list of words and their respective definitions that explain difficult, technical, product-specific, company-specific, or industry-specific terms. The glossary will help guide the linguist in understanding the meaning and context for each of the specialized terms so that they can, in turn, select the best translation. In most cases, just the terms are translated, not the definitions (unless they appear in the original document).

Glossary terms are selected from input gathered from technical writers, software engineers, marketing departments, etc. Your provider may also extract likely terms from source content that has been provided to them. The list is reviewed by client representatives and subsequently translated. The approved list of translated glossary terms forms the basis of the terminology list. Translation quality and project costs benefit from the creation and maintenance of a glossary.

Terminology List

A terminology list (and its translations) contains the most important terms from the original source materials. These terms may be derived from a variety of sources:

- The glossary developed by the technical writer of the source document,
- Selected operating environment terminology of major software developers (e.g., Microsoft),
- Software and documentation that the client may have previously localized,
- Other localized resource materials such as marketing collateral and product lists,
- Company standards (such as part numbers, technical and product support information, warranties, license agreements, copyrights, references to other software programs, product names, brand names, and non-translated terms), and
- Country standards for expressing functional or cultural dictates such as publishing standards, sorting of lists, abbreviations, time, dates, holidays, currency, and measurements.

There are several ways to create a terminology list. You, or someone in your company, may put it together based on expertise with the material. If you opt to have your vendor compile the list, they can use either a manual process or one of several automated tools to identify the key terms. Another option is to have a linguist develop and translate the list simultaneously. Take caution, however, when doing this for more than one language. If separate linguists are creating lists for several languages simultaneously, chances are you will end up with terminology lists of varying quality and completeness.

Regardless of how the terminology list is developed, it needs to be reviewed and approved by a company representative before being used. The list is only as good as the review and approval process, and all stakeholders need to be in full agreement regarding the terms and translations on the list prior to the project start. If a dispute arises over the translation of a key term later in the project, it could have a severe impact on budget, timeline, and of course, quality.

Once approved, the terminology list becomes the basis of the initial translation memory (see Chapter 9 for a detailed discussion of TMs) and is used by all linguists involved in the translation phase. If any additions, deletions, or modifications to the terminology list are necessary, it is important that, once again, all invested parties sign off on them. Remember, since this list contains the key terminology for your project; be certain everyone agrees it is right.

Carlota Ribas-López
Senior Project Manager

Born into a bilingual household in Barcelona (Spain), I have always been aware of the power of language and culture. Working in the localization industry, where we are exposed daily to multiple languages and cultures, feels very comfortable to me.

What if you decide not to develop a terminology list? In some cases, that is the right decision for the level of quality required on a project. If you are only translating an email for comprehension, then a terminology list is probably not needed. But if this is your corporate website, instructions for how to install and use your new product, or training steps for new SAP procedures, then not developing a terminology list could be a costly and critical mistake.

However, preparing a terminology list does not mean that your linguist can simply accept the approved translation each time the key term appears. The translation of certain terms can vary depending on the context (e.g., whether they refer to software or hardware, or whether they are being used in a formal, informal, or imperative form). Further, some terms do not translate well and are better left in English; other terms are better abbreviated. Linguists need to use care in applying the approved terminology list, as do you and your representatives.

A terminology list is not just another way for your vendor to charge for additional services. A terminology list forms the backbone of your translation projects, directly contributing to your project's consistency and quality—from your first project to the ones months and years down the road. You should consider the pros and cons of having or not having a terminology list and be ready to build into your schedule the necessary time and budget to develop one.

Some of the direct, quantifiable gains in quality that can be achieved by using a terminology list include:

- The translator, copyeditor, and proofreader all use the same industry-specific terminology throughout the project and over all project components,
- Consistency of abbreviations, product names, non-translated terms, and measurements,
- Consistency between country and company standards,
- Local (i.e., in-country) suitability, and
- Consensus among client, distributor, and localization provider.

oops

At the Beijing Airport:
Please check in animals and alcoholics

Examples in Spanish

Agreement on terminology
"Congratulations" can be translated correctly as **Felicitaciones** or **Enhorabuena**.

Local suitability
"Congratulations!" as well as "Welcome to ..." are frequently used in user manuals to introduce a new product. Should the Spanish audience be addressed in this rather colloquial American way? Is there a more formal way to address the user, or should this greeting not be used at all?

Abbreviation
In all cases the abbreviation is written first, with the name for which it stands written in parentheses. However, there appears to be no set standard on the placement of the translated text. Client and localization vendor need to agree if the translated text should be placed immediately after the abbreviation or after the name for which it stands.

UK — **Reino Unido** (United Kingdom)

Examples in Japanese

Depending on the platform, commands and buttons are translated differently:

Save As	別名保存	名前を付けて保存
Cut	切り取り	カット
Print	印刷	プリント

Depending on the context, an English word can be translated into multiple terms in the target language:

Address	アドレス、住所
Title	題名、タイトル、呼称
Class	クラス、級、レベル
Time	時間、タイム

On the other hand, sometimes multiple terms in English can be translated into a single term in Japanese:

Tall	高い
High	高い
Expensive	高い
Pretensious	高い

Some words and abbreviations, by convention, stay in English:

lpi	**lines per inch**
pts	**points**
m/cm/mm	**meter/centimeter/millimeter**
g/kg/mg	**gram/kilogram/milligram**

Leah Carter
Accounting Manager

1 is the number of siblings I have.

2 is the number of positions I've held at Lingo.

3 is the number of kids I have.

4 is the number of states I've been in.

5 is the number of other countries I've visited.

6 is the number of years I've been with Lingo.

7 is the number …

I like numbers, can you tell?

Examples in German

Variation between software and hardware technology
"Setup" is translated into **Einrichten** if the term refers to setting up the software, and **Anschließen** if the term refers to setting up a peripheral device.

Non-translated term
In projects where the documentation is translated but the user interface stays in English, there should be an agreement about whether the English term is followed by the localized term in parentheses or vice-versa.

Klicken Sie auf **Load/Unload Panel** (Stück laden/Entfernen), or

Klicken Sie auf Stück laden/Entfernen (**Load/Unload Panel**)

Style
"Connect your printer to the computer" can be translated formally into:

Schließen Sie den Drucker an den Computer an.

Or in the imperative voice:

Drucker an den Computer anschließen.

Or in the passive voice:

Der Drucker muss an den Computer angeschlossen werden.

Linguist Selection

Professional translations require professional linguists. It is really that simple. On an abstract level, most people intuitively understand that a professional linguist will produce higher-quality translations than an employee who happens to speak the language or part-time college students trying to earn a few bucks on the side. They also understand that, as with all professions, not all linguists have equal skills or subject matter expertise. Nevertheless, the allure of saving a few dollars is hard to ignore. Time and again, a native Japanese- or Spanish-speaking employee, who may not even be familiar with the product, is asked for help in translating it.

In developing the English source, you probably used professional technical writers to ensure that the English content was correct and appropriate for the end user. For precisely the same reasons, it also makes sense to hire professional linguists to translate these materials. Engineers should do what they do best: design software, bridges, cars, and staplers. Most engineers are *not* technical writers. So, if you would not consider asking your software engineer to write your English user guide, why would you want to use a software engineer as a translator? Translators are highly trained professionals who know how to get your message across clearly and correctly, so that your end users can easily understand the information.

Localization involves much more than word-for-word "translation." Because different cultures use different grammar and sentence structures, straight word-for-word translations often do not reflect proper linguistic style or accurately capture complex concepts. Sometimes they no longer even convey the original meaning or message. An experienced linguist, however, can accurately translate the most difficult materials and also convey nuances, ideas, and register (tone, style, formality, complexity, etc.).

Localization quality is directly linked to the translator's experience with the topic and knowledge of both the source and target languages. The translators working on your project should provide native-quality work. Native quality means that the material, once translated, reads as though it was originally written in the target language. This usually requires the expertise of someone raised and educated in the target country. Of course, there are non-native translators with exceptional education, training, and experience in a specific language who are able to provide native-quality work, but that takes true talent.

So just what makes good a translator? A professional linguist will possess:

- Native fluency in the target language,
- A thorough understanding of the source language,
- Excellent writing skills, including a grammatical mastery of the target language and knowledge of various written forms and styles,
- Familiarity with current terminology in the desired field (experienced translators maintain extensive reference libraries),
- A working knowledge of the localization process,
- Access to appropriate tools, such as up-to-date computers, multiple software applications, and industry-specific software tools such as Trados, and
- An acute awareness of cultural differences and language subtleties.

To obtain the highest-quality translations, your localization vendor should have well-documented linguistic qualification procedures (and follow them!) for evaluating and hiring individual linguists or translation agencies. The linguists they use for your projects should both be experienced translators and be familiar with your content subject matter. There are various means to assess competency, and no single method is necessarily the "right" way. However, to ensure the highest quality it is imperative that your localization vendor has procedures in place to verify that only qualified professionals with appropriate subject matter expertise are assigned to your project.

Ann Chay
Business Development
Manager

Strong family ties were taught to me by my grandparents who migrated to Oregon from Italy at the age of 6. I was raised in a small suburb, Milwaukie, with all my aunts, uncles and cousins. I graduated from Portland State University and decided to move to San Jose, CA. After living in California for 8 years my husband and I decided to move our family back to Oregon. My husband is Korean, a culture where family ties are also very strong. So our boys are being raised with the double whammy of Italian and Korean family values.

At Lingo Systems, we believe that a translator should meet the following clearly defined minimum criteria:

- A Bachelor's or Master's degree in an appropriate field,
- Five years of general translation experience,
- Three years of translation experience with material similar to the source material,
- Translation certifications such as those provided by the American Translators Association, and
- A demonstrated commitment to the profession through professional affiliations.

The quality of your final deliverable depends on the professional abilities of many individuals at every phase of your project's development: software Help file engineering, linguistic QA, desktop publishing, etc. Keep in mind that each one of these discrete tasks requires individuals who are localization professionals. Using qualified and experienced translators will infuse your products with a professional style and clarity of content that contributes to the success of your international release.

A Three-Step Translation Process

Even the very best linguists are not perfect. Thus, another important factor in achieving the highest quality is the number of linguistic steps that your vendor performs. Many vendors use only one or two linguistic steps to localize source materials. At a minimum, a basic translation step is always performed. A second copyedit step—hopefully by a second professional linguist—is often employed as well. To provide the highest quality, however, a three-step process, with a different linguist performing each step, is the best way to ensure an accurate translation. At Lingo Systems, unless instructed otherwise by our client, we perform three separate linguistic steps, each by a different linguist:

- Translator,
- Copyeditor, and
- Proofreader.

The translator is the "lead linguist" on your project and is responsible for converting the source material content into the target language. The copyeditor then reviews, word for word, the lead translator's work, verifying the accuracy of the translation, including double checking it against the glossary, terminology list, or style guidelines that were developed at project start. Lastly, the proofreader examines the copyedited version on a "standalone" basis for consistency, proper format, and flow of the language.

While most commercial products should use a three-step linguistic process, there are cases where this is "overkill." If you simply need to have your content understood in the target language, one translation step, or translation followed by either copyediting or by proofreading, may be sufficient. A two-step approach is particularly useful for internal documents like training materials and memos.

Your vendor should be flexible in assembling a translation team that best satisfies your budget, timeline, and quality requirements. For example, you may have a 40,000 word project that you need in three weeks. The average number of words a professional linguist can translate in one week is 10,000, so clearly your deadline cannot be met if only one translator is assigned to the project. Instead, a team is needed that consists of three or four translators working on different sections of the project, two or more copyeditors, and one proofreader who, in addition to ensuring consistency of terminology would also make sure that all translated material reads as one voice. In this case, given the large number of words and a short timeline, you can see that the development of a glossary, a terminology list, and a style guide would go a long way toward providing a high-quality deliverable!

There are many possible linguistic team configurations that your vendor can use to successfully deliver a high-quality product. Sometimes, a team of multiple translators, copyeditors, and one proofreader make the most sense. On another project a single translator, multiple copyeditors, and multiple proofreaders works best. Your vendor might also use multiple translators and then have them copyedit one another's work before sending the content on to the proofreader. The key to quality when multiple linguists are working on your project is having your vendor develop a solution that ensures that the final product reads as one voice and that terminology is applied consistently. Usually that means that one linguist on the team— whether it is the translator, copyeditor, or proofreader—will review the content of the entire project.

The bottom line is that flexibility and a little creativity, along with careful preparation, will serve you well on virtually any project no matter how big the word count and how tight the deadline.

In-Country Review

The final procedure that should be employed to maximize quality is an in-country review. Check the quality of your vendor's work by using your own in-country representatives (employees, distributors, customers, or agents) to review the translation. This is an excellent way for you to be sure that all materials were consistently translated, and that your message has been accurately reflected. Whomever you select, the best results will be obtained if the reviewer is familiar with your products. These reviewers can then evaluate the specifications of your product against the cultural and linguistic elements of the relevant country.

When developing an in-country review process, keep the following in mind: If possible, involve your in-country reviewers at the glossary development stage, so that they are aware of all the terminology choices; bringing an in-country reviewer on board at the project's start avoids costly revisions that may arise due to stylistic differences in the translation later in the game.

Ben Solis
Project Manager

I firmly believe that cultures that can communicate with one another can find some common ground for peace and prosperity. I don't think it is a far stretch to suggest that by facilitating communication, we're creating a better world.

In addition, be sure to appoint a single chief reviewer to collate and approve all of the change requests for each language; this avoids the potential for mutually contradictory changes from different reviewers with different opinions as to the style. Your localization vendor can make any change in the translation that you request, but you must be clear in communicating the change that you want.

The materials you entrust to your localization vendor represent your company to your customers. It is how they meet and interact with you. Quality has to be more than a catchphrase used to sell a service. It has to be at the core of the localization process. The quality of your final deliverable depends on the professional abilities of many individuals at every phase of your project. By utilizing localization professionals who follow the steps and procedures described in this chapter, you will gain an edge on your competition, maximize your investment, and maybe even sleep better at night knowing that your original product looks, feels, and sounds just as it was intended for each of your target markets, which is as it should be.

oops

Backward market rates:

Estimates of growth in the forward market rates in China were mistranslated by the official Chinese translators without the subjunctive case (the case of possibility) and then reported worldwide by Bloomberg. The mistranslation stated that China had decided to revalue the yuan by 1.26 percent within a month and 6.03 percent in 12 months. Such a revaluation would force a revaluation of all Asian currency, and nearly 2 billion dollars were traded over a few minutes time before the error was discovered and Bloomberg reported the error.

Chapter 8
Same Language, Different Dialect

The difference between a dialect and a language could just as well be a geographic or even political distinction as much as a linguistic one. For example, a Dane, a Swede, and a Norwegian can speak their own language at a party and understand each other. It might be a little harder for them to understand an Icelander, but more or less, the guests will still get the blonde jokes—well, maybe not the Swedes. This is because all of these Scandinavian languages have common roots from Old Norse. However, it is a sensitive issue to discuss whether Norwegian is a dialect of Swedish, if it is the other way around, or to identify a truly pure Scandinavian language. Therefore, it makes sense to call all of these varieties languages instead of dialects, and to separate the dictionaries.

From an anthropological standpoint, however, everyone speaks a dialect of some other language. Winston Churchill once described the US and the UK as two countries separated by a common language. That description extends to other languages as well. When translating written material in any form, it is critical to choose the right dialect in order to maximize the return on your localization investment. Some of the more common languages where these very important distinctions arise are French, Spanish, Portuguese, and Chinese.

Spanish (Iberian vs. Latin)

Spanish localization represents one of the most obvious opportunities to expand the market for your products. More than 410 million people around the world use Spanish as their native language. In the US alone, roughly 35 million residents primarily speak Spanish, and that number is expected to grow substantially in the coming years.

The fact that Spanish is spoken so widely poses an interesting challenge for localization because the language has evolved in each region of the world. Spanish translators typically make a distinction between Iberian (Spain) and Latin American (the Americas) Spanish. The differences are further compounded, however, when examined on a country-by-country basis. Although the broad brush of Latin American Spanish includes Argentina, Colombia, Mexico, Cuba, and Puerto Rico (to name a few), there are clear linguistic variations and peculiarities that characterize each, with the greatest being pronunciation.

The main stabilizing force for this highly dynamic language is the Real Academia de la Lengua Española (Royal Academy of the Spanish

Ryan Kristin
Account Manager

Every since I began traveling internationally years ago I have been fascinated with new and unfamiliar cultures. At Lingo I have the privilege of communicating with people all over the world on a daily basis. It is exciting to work in a place where all languages, races, and customs come together for a unified purpose.

Language). This language academy oversees changes in grammar and vocabulary for all Spanish-speaking countries, including those in Latin America. Their decisions are meticulously observed by those who teach, write, or are in any way involved with the use and implementation of the Spanish language.

Given the lasting impact of their decisions, the Royal Academy is painstakingly slow in reaching those decisions. While the arbiters of new terminology may proceed at a very slow and cautious rate, technology races along. Until the Royal Academy decides each issue, the Spanish translator is forced to make his or her own decisions on terminology. Experienced translators are always careful to use terms that are understood by the greatest number of users, regardless of their location.

If Spanish has a regulating body that decides all matters concerning written Spanish (grammar, syntax, spelling, etc.), then why do some people believe that there are different kinds of Spanish? As noted above, the greatest differences exist in the way words are spoken and the way certain letters are pronounced (or maybe not pronounced). Thus, in certain parts of Spain the letter "z" is pronounced as a soft English "th" as in the word "thin," whereas in Latin America, the letter "z" is always pronounced as an "s" as in "Sam." However, whether in Argentina, Mexico, or Madrid, the word "zapato" (shoe) must always be written with a "z." Local differences can also be found in the use of certain nouns, especially those that designate agricultural products: The English say "potato," Latin America prefers "papa," and Spain "patata."

Geography can also play a role in the determination of terminology. With their geographical proximity to the United States, some Latin American countries identify more closely with terms used in the United States and "Spanish-ize" the terms. A good example is the word "computer." In most Latin American countries "computer" is rendered as "computadora." In Spain, because of its proximity to France, "computer" is rendered as "ordenador," from the French "ordinateur." However, geographical proximity is not always the determining factor; take, for example, the English term "font." Latin America prefers "tipo" or "fuente," while Spain has kept the English word "font."

While these examples contrast Iberian and Latin American Spanish, other linguistic differences occur within Latin America. Chile, Colombia, and Argentina usually identify more closely with Europe than the United States, yet the rule is not hard and fast. The decimal and thousand separators are good examples. Mexico and Central America use these separators in the same way as the United States (where one thousand twenty is represented 1,020.00). South American countries prefer the European way of expressing separators (where one thousand twenty is represented 1.020,00).

In the United States, the situation is even more complicated as Spanish speakers have emigrated from many different countries and brought their regionalisms with them. How, then, can you choose the best "regional" form of Spanish to translate into?

This is the key question that you must answer once you conclude that there is a business case for making your materials available to a Spanish speaking audience. At a minimum, you must decide between Iberian and Latin Spanish as no one "flavor" will always be fully suitable. In our experience, the differences are sufficiently significant that most companies will elect to do both when trying to reach a worldwide Spanish speaking audience. If you want to minimize your localization investment but still reach as broad an audience as possible, however, "International" Spanish may be an adequate compromise.

Alternatively, if you must reach a very specific audience, such as Spanish for Colombia, your localization partner should provide experienced linguists who can do that. Select the market where the majority of your audience resides, or select two of the major regional forms. When setting out to translate into Spanish, you must identify your target audience and then select the regional form(s) to best reach them. Your localization partner can help you with this decision and then provide Spanish translators with extensive experience to ensure your materials are best understood by your target readers.

Portuguese (Portugal vs. Brazil)

Nearly 210 million people speak Portuguese throughout the world. However, spoken Portuguese is not homogeneous. It differs in grammar, pronunciation, and vocabulary among Portuguese speakers in Portugal and in Brazil. Moreover, unlike Spanish where a "neutral" or "international" Spanish may be adequate for most markets, it is very rare for companies to translate their materials into one type of Portuguese for both markets.

Brazilian Portuguese was not only influenced by native languages such as Tupinambá, but also by the many languages spoken by African slaves. Although some Brazilian words made their way to Europe, most were only used in Brazil. Southern Brazil absorbed a large influx of immigrants of Italian, German, and Japanese descent. These linguistic groups made several contributions to the language spoken in Brazil. Portuguese in Europe, meanwhile, was influenced by the French spoken during Napoleon's occupation of Portugal.

In the 20th century, the linguistic split between Portuguese and Brazilian increased as the result of technological innovations that required new vocabulary. Unlike the Royal Academy of the Spanish Language, there

is no similar "watch dog" to approve new terminology and grammar in Portuguese, and these two languages are increasingly growing apart. Brazilian Portuguese tends to be a more flexible language, much influenced by North American culture. For instance, Brazilian Portuguese translates the word "user" as "usuário" (a nonexistent word in European Portuguese), whereas European Portuguese translates it as "utilizador," similar to the French "utilisateur."

Internet World Magazine published a list in the Brazilian edition that pointed out some additional differences:

English	Brazilian Portuguese	European Portuguese
to access	acessar	aceder
mouse	mouse	rato
screen	tela	ecrã

Besides words that are completely different and/or are used in a completely different context between these two variants of Portuguese, there are approximately 400 words with a different spelling and 1,500 with a different accent mark. There are also some grammatical differences.

When localizing into Portuguese, be aware of these differences. Your localization provider should distinguish between European and Brazilian Portuguese and should use native-quality speakers from the appropriate country to localize your product. Although Portuguese speakers from both sides of the Atlantic can usually understand each other, not localizing separately for each market can lead to confusion among your end-users.

French (France vs. Canada)

There are now nearly seven million French speakers in Canada, mostly located in the province of Quebec. Over the past 400 years, the French spoken in this region has evolved dramatically, due in large part to its distance from French speakers in Europe and its close proximity to English speakers in Canada and the United States.

Canada has certainly become a prime target market for localized products. In 1988, the Canadian government passed the Official Languages Act, which conferred equal "official language" status to both English and French. This was done in order to preserve the nation's French linguistic heritage and to "support the development of English and French linguistic minority communities and generally advance the equality of status and use of the English and French languages within Canadian society." As a result, all official federal government communications must be made in both languages, and government services must be available in both languages. Many commercial products follow the government's lead and provide packaging, labeling, etc., in both English and French.

Jordan Bulloff
Project Manager

Growing up wrestling crocodiles in my backyard in Australia, fighting sheep in England, and commonly provoking bears by eating honey in front of them in Canada, I quickly learned the importance of foreign language skills to stay alive. Learning popular survival phrases such as:

"Crikey Croc, don't eat me … have some Vegemite instead."

"Oy Mate, leave me alone … B'aaaah!"

And my personal favorite … "I can't outrun the bear, but I can outrun you, ey?" Are the only reasons I'm still here today.

Only the strong survive … Only the strong survive …

However, if you provide European French to a Canadian audience, you may be missing the mark. Canadian and European French differ in many ways, including vocabulary. Some differences include the following:

English	Canadian French	European French
telephone handset	le récepteur	le combiné
You're welcome.	Bienvenue	De rien (and several others)
blueberry	le bleuet	la myrtille
soccer	le soccer	le football
snowbank	le banc de neige	la congère
go shopping	magasiner	faire des courses

In addition, there are differences in pronunciation, in pronouns, in verbs, and in usage.

What, then, should you know when deciding whether to translate solely into French or into both French and French Canadian? French Canadians understand any material translated in French, since the written language is generally similar. If simple understanding is your goal, the expense of translating into both French and French Canadian may not be necessary.

However, if you want French Canadians to feel that your product has been custom-made for them, you should translate it into French Canadian as well to ensure cultural sensitivity. This can usually be accomplished by having the product of a French translator copyedited by a Canadian colleague. Most of the time the linguistic changes are minimal, but you can then be confident that your product is indeed targeted for Canada.

Chinese (Mandarin vs. Cantonese, Traditional vs. Simplified)

"Can you speak and write Chinese?" This apparently simple question can be answered by asking in turn, "Do you mean Mandarin, Hakka, Cantonese, Traditional Chinese, Simplified Chinese, or ...?" It seems that there is quite a bit of confusion regarding what exactly "Chinese" means in regard to both the spoken language and the written language. Let's try to clear up some of this confusion.

First of all, spoken Chinese consists of dozens of different dialects, often mutually unintelligible from one another. Pu Tong Hua, known as "Mandarin" in most Western countries, and Cantonese are the two most widely spoken of these dialects. As early as the second century BC (during the Han dynasty), language reforms were implemented in an attempt to standardize the language. Mandarin evolved as the official Chinese spoken language because it was derived from the Beijing (i.e., Peking) dialect,

Vanessa Wang
Project Manager

I grew up in Taiwan, an island thousands of miles away from Portland. It's the passion for communication that brings me here. The challenges of unpacking and conveying messages to people speaking different languages make this industry fascinating, and being involved with people from different cultures adds to the fun part of working at Lingo.

which was taught by scholars and used by the government for nearly 1,000 years. Taiwan and Singapore also use Mandarin as their official language. Cantonese, on the other hand, is a dialect widely spoken in the southern regions of China (the Guangzhou and Hong Kong areas).

Mandarin and Cantonese are the most widely spoken Chinese dialects, but they are by no means the only ones. For day-to-day conversation, many people still prefer to speak in the dialect of their respective regions. It is common to find that two people speaking two different dialects cannot communicate verbally; however, they can communicate in writing thanks to the standardization of the Traditional and Simplified writing systems.

The origin of the Chinese writing system is pictorial, dating back thousands of years. People drew pictures to express their thoughts—in short, to communicate. As you can imagine, this method of written communication was very cumbersome, making complex thoughts difficult to express. As a result, a number of reforms have been initiated to stylize and simplify the manner of writing Chinese. This has, in turn, resulted in a more uniform writing style.

Of all the language reforms initiated over the past two millennia, none has had a greater impact than the one carried out by the People's Republic of China (PRC) after its establishment in 1949. The mid-20th century language reform simplified the characters used in the Traditional Chinese writing system by reducing the number of strokes needed to write a character. The end result was the Simplified Chinese writing system. The PRC and Singapore currently use the Simplified Chinese writing system, while Hong Kong and Taiwan use the Traditional Chinese writing system. However, now that Hong Kong has been integrated into the PRC, we may see an increase in the use of Simplified Chinese there.

It is generally easier for a person who knows Traditional Chinese to understand Simplified Chinese characters than a person who knows Simplified Chinese to understand Traditional Chinese characters. However, this is not a sure thing. Moreover, using one of the standard writing systems is still not always sufficient for proper localization. For example, the Traditional Chinese used in Taiwan is somewhat different from that used in Hong Kong. The character set is the same, but they are sometimes grouped differently to express various concepts. One example of this is the term "lunchbox." In Hong Kong, this term is expressed by pairing the character for "rice" with the character for "box"; in contrast, Taiwanese use a pair of characters that roughly translate as "convenience now." A Hong Kong reader unaware of Taiwanese culture would understand the characters but not grasp the lunchbox connotation. If you are planning to localize for the Hong Kong market or for the Singapore market, you should plan a special copyediting step in order to customize your translations for those specific markets.

Continuous efforts at language reform introduced the use of the Roman alphabet to "spell" the pronunciation of Chinese characters. The result was the standard Pin Yin spelling system that is widely used in China, Taiwan, and Singapore today.

So, what is Chinese? You can see now that it really is a "blanket term" for several major dialects and two major writing systems.

Country	Spoken Language	Written Language
PRC	Mandarin	Simplified Chinese
PRC Guangzhou Province	Cantonese	Simplified Chinese
PRC Hong Kong	Cantonese	Traditional Chinese
Singapore	Mandarin	Simplified Chinese
Taiwan	Mandarin	Traditional Chinese

The next time you hear some one say, "We need this translated into Chinese!" you should first think about what the term "Chinese" really means!

As the means of communication becomes more efficient and accessible to people around the world, one has to wonder what the effect will be on language. Will languages become more homogeneous or will its users create their own identity by supporting distinct dialects? Will language differences be based increasingly on economic or age-related factors, geographic boundaries, relations with other countries, or possibly ideology? Perhaps a cultural identity will prevail over a regional one, where physical distance might not have any influence at all. However these factors combine to influence language, it will have a direct impact on your business. Having a trusted localization partner, like Lingo Systems, to advise and assist you will be critical.

oops

In an East African newspaper:

A new swimming pool is rapidly taking shape since the contractors have thrown in the bulk of their workers.

Cédric Vezinet
Director of Engineering

Cédric started working at Lingo Systems in 1996. Even though he is shy and a bit of an introvert, he is always happy to help around the office. He has some localization knowledge which comes in handy once in a while. But overall we just keep him around to tease him and his French heritage. Our favorite time at the office is when we hang him by his feet on Cinco de Mayo and use him as a piñata.

Chapter 9
Engineering and Computer-Aided Translation Tools

Localization is a multi-disciplinary activity that includes linguistics, formatting, and quality assurance. Engineering is also an integral component of this service and is one of the tasks that differentiates simple translation from comprehensive localization.

Localization engineers are involved at every stage of the localization process. Often they consult on internationalization matters before the materials are even developed. Once source files are created, the engineers' analysis of them provides vital information for project planning and budget estimation. Linguists rely on engineers to extract or isolate text strings from source content and prepare marked-up files to facilitate translation. Localization engineers also manage the ensuing translation memory and use tools such as Trados, Passolo, Catalyst, and Multilizer to improve consistency and lower costs. Prior to delivery, engineers may perform the functional and/or localization testing of the localized products.

At Lingo Systems, members of our engineering group closely interact with the other production departments to provide further support. For our formatting team, they import and export text from DTP applications. For our QA department, they perform functional testing of technical projects such as user interfaces, websites, multimedia content, and help systems. And they are always available for a quick game of Wii bowling over lunch.

First Things First: Internationalization

Many US- and UK-based companies develop their products with only English-speaking customers in mind. When these domestic products are slated for distribution to foreign markets, the process of localization often reveals limitations in the product design. Internationalization is the process of engineering a product so that it can be localized for export to any country.

Often, internationalization is quite simple. For example, some languages use more characters and take up more space than others. A properly internationalized source file will leave room for text expansion or its user interface will dynamically resize itself. Another common internationalization step is to resize an 8 1/2" x 11" document to European A4 paper size.

In addition to considering overall design and layout, the internationalization process focuses on several other important issues:

- Does the design account for cultural differences in various metrics such as currency, units of measure, date format, phone numbers, and addresses?
- Are all the localizable strings externalized or isolated from variables and other code for easy extraction?
- Are unique strings reused in different contexts throughout the product?
- Is the product free of embedded and concatenated strings?
- Do automated lists take into account any sorting order differences in the target locale?

To avoid internationalization surprises, involve your localization provider during the product design stage so that localization requirements can be taken into consideration during development. If this is not done, your localization vendor will likely have to perform some product internationalization prior to beginning localization. This may not only compromise timelines, but may also have an adverse effect on your budget.

Encoding: Pick Your Poison

A major question to address when you begin the internationalization process is: Can or will your application use some version of Unicode as its encoding format? Before Unicode was invented, there were dozens of different encoding systems. No single one contained enough characters to represent every possible language. For example, the European Union alone required several different encodings just to cover its languages. Even for a single language like English, no single encoding was adequate for all the letters, punctuation, and technical symbols in common use.

To add to the challenge, many of these encoding systems also conflict with one another. That is, two encodings will use the same numeric assignment for two different characters, or use different numeric assignments for the same character. Computers (especially servers) must be able to support many different encodings, but it still may not be enough. Whenever data is passed between different encodings or platforms, it runs the risk of being corrupted.

Unicode eliminates most of these problems. It is well established, works on all platforms, and supports many more characters than most of us have ever heard of or will ever use. Unicode provides a unique number for every character, no matter what the platform, no matter what the program, no matter what the language. It also allows data to be transported between many different systems without corruption. Due to the natural progression of technology, there are many different Unicode formats: big-endian, UTF-7, UTF-8, UTF-16, UTF-32, and so on. In general, UTF-8 will be

Christophe Vezinet
Senior Localization Engineer

Iznogoud (pronounced "is no good" with a French accent) is a French comic series featuring an eponymous character. Iznogoud is the second in command (Grand Vizier) to the Caliph of Baghdad, Haroun El Poussah—but his sole aim in life is to overthrow the Caliph and take his place. His famous catchphrase, "I want to be Caliph instead of the Caliph" ("je veux devenir calife à la place du calife"), has passed into everyday French for qualifying over-ambitious people who want to become chief. Like Iznogoud, I slowly make my way in the difficult world of Localization Engineering to become one day not the Caliph but the Napoleon of Localization. (See *The Guide to Translation and Localization, 6th edition*). I am getting closer every year, to be continued …

most common on the Web. UTF-16, UTF-16LE, UTF-16BE are mostly used by Java and Windows. UTF-32, UTF-32LE, UTF-32BE are mostly used by various UNIX systems. Fortunately, the conversions between all of them are algorithmically based and quick to implement.

Pick Early, Test Often

Many internationalization issues can be identified early in the development process by performing internationalization testing of the source material. Machine translation (MT) technology is often used for this purpose since it can generate pseudo-translated content that has the characteristics of translated material without a costly investment in translation. Machine translation, which is described more fully in Chapter 20, is based on advanced computational linguistic analysis and, because it is cheap, can quickly generate lots of translated content for testing purposes. Such testing can help pinpoint issues in the localization project before they become major headaches. For instance, a pseudo-translation can identify variables in the software that should not be translated, allowing you to isolate them prior to actual linguistic work.

Internationalization is not a service commonly offered by localization companies as it requires highly skilled and specialized personnel with a very strong understanding of the platforms and development environments being used. Moreover, a well-executed internationalization review will not necessarily rid your files of all potential localization headaches—but it will reduce them to a manageable level and avoid the introduction of additional defects during the localization process.

In general, the difference between a successful project and one plagued by problems is a direct function of the amount of interaction between the client's and the vendor's engineering departments in the early stages of the project. Internationalization evaluation and testing is a very cost-effective way to ensure that your product is ready for localization—especially when measured against the delays and costs associated with trying to resolve these issues during the localization process.

On Our Way: Localization Begins

Once all internationalization issues have been addressed, the localization process can begin on a good foundation. For the engineering group, this usually means preparing the source files for translation. How this is done varies depending on the type of materials. The five main categories are:

- Documentation localization,
- Help localization,
- User interface (UI) localization,
- Web localization, and
- Multimedia localization (subtitling, audio/video recording, etc.).

Page by Page: Documentation

Want to see us pull a rabbit out of our hat? Well, perhaps that's a stretch, but this is where the magic starts. Imagine you have just purchased the latest and greatest Widget. The first thing you do is read the manual, right? (C'mon, work with us here.) Now, imagine lifting all the English out of that manual, crunching it all up and then carefully unfolding it to reveal a brand new language. It's a bit of a strange notion, but that's pretty much how document localization works. In the simplest terms, documentation engineering is the process of importing and exporting text from a desktop publishing application or content management system (CMS).

Since most translators work within Microsoft Word using computer-aided translation (CAT) software, the source material (which can be in any medium) must be converted to an RTF or TTX file while preserving the formatting of the original document for the linguist to be able to work. This is done by using different tagged text formats to isolate the formatting from the translatable text. By protecting the formatting, the translators can then use their CAT tools and focus exclusively on what needs to be translated without being confused by formatting codes, which can be very numerous (especially in the case of Quark documentation).

The vast majority of documentation is developed using Adobe InDesign, Adobe FrameMaker, QuarkXPress, and Adobe PageMaker. As you can see, Adobe Systems Inc. has quite a few different writing tools, but as time goes by they seem to be moving toward one versatile application that will address all documentation development needs. Adobe ceased production of PageMaker in 2004, replacing it with InDesign, which continues to make the localization process easier. InDesign is a terrific application for localization, as it offers full Unicode support and is well-suited for cross-platform work. It also allows for XML integration with a CMS. But we must admit that there are quite a few deficiencies in the INX (InDesign Exchange) and Tagged Text format which makes the localization process a little bit tricky. But we always have a bit of engineering magic up our sleeves to solve the problem.

Regardless of the application used to develop the materials, when the RTF or TTX files come back from translation they head straight to the engineering group. With a wave of a wand and some feverish keyboard tapping, engineers pour the localized text back into the source documents and hand them off to the DTP department where they are polished to perfection.

Chris van Grunsven
Senior Localization Engineer

Much has changed since Lingo started. Windows was an MSDOS program. You had to write your own software if you wanted a multimedia presentation. And linguists had bookshelves full of previous translations to match the style of various clients. Now linguists have tools to automatically reuse previous translations on their computers. We work with voice talents for multimedia projects and videos. And Linux makes up a significant part of our jobs.

Stop the Presses: Help File Localization

As a means of disseminating information, print documentation is quickly losing ground to interactive help systems. Well-structured online help provides users with incredible search capability, allowing them to find more information in less time than with conventional print documentation. Many help users say this leads to a richer experience. We could not agree more.

Help systems are not only getting bigger, they are getting smarter. Perhaps most importantly, however, they are becoming easier and less costly (if not downright cheap) to create. Single source publishing tools such as Author-it, Arbortext, WebWorks, or RoboHelp are now able to import previously generated Word or FrameMaker documents and then leverage them to create interactive help systems. Let's note here that localization-savvy applications which offer built-in localization support, such as Author-it, make the translation process a walk in the park. As more companies discover these benefits, this trend will only accelerate.

The main help formats we see being used are WinHelp, HTML Help, WebHelp, JavaHelp, Oracle Help, and FlashHelp. Even though all these formats have their own specific uses, when it comes to localizing help systems, the approach is similar.

Interface This: Software Localization

An engineering group really shines during the localization of software. We take on all comers: any version of Windows, Mac OS, UNIX, Linux, Palm OS, Symbian, mainframe, and Java-based applications. And we will take any variety: web-based, server-based, or client-based.

For some programming language and platform combinations, software localization requires a process not unlike the one used for documentation. The localization engineer extracts the text from the application and then creates a tagged RTF or TTX file for the translator that protects the underlying codes. When it comes to protecting the codes, TTX is by far the better choice. In other cases, the localization engineer uses a proprietary tool or an off-the-shelf application like Catalyst, Passolo, or Multilizer that allows the translator to work directly on compiled files and executables. All things being equal, however, it is more common and easier to work in the resource (RC) or properties files to minimize the amount of preparation and reduce the potential for defects being introduced during the localization process.

Whatever method is used, one thing is sure: The continuing evolution of Unicode technology and the greater understanding of the needs of the international market have made localization engineering much easier. The latest OS editions from Microsoft and Apple are perfect illustrations.

Since Windows XP and Office 2003, dealing with multiple languages in your day-to-day operations has become much easier. It is now possible to easily generate text files in many encodings for the most widely used languages on any Western operating system. The manipulation of Eastern languages, double-byte, and even right-to-left languages has been made much easier as well. We previously had to navigate from one native operating system to the next just to manipulate localized files. Much of this hassle has now disappeared and native operating systems are only used for online functional testing of the final localized product.

Also widely used and indispensable is Apple and its Mac OS X, especially its latest Leopard release (and soon to come Snow Leopard). Not only is it a great system for localization but, in our opinion, the most localization-friendly operating system on the market. With just a simple drag of the mouse, users are able to switch the UI and/or the system's language! And the management of multilingual resource files is the best we have seen.

No matter what the platform, the best way to make your UI localization-friendly is to externalize all localizable strings (similar to Java's properties file). Whenever possible, design your UI so that most of the strings are located in well-formatted files where the variables are followed by the string and the interface is dynamically laid out. Another important rule is to avoid string concatenation at all cost.

Going World Wide: Web Localization

User interfaces are increasingly web-based because they are easier to maintain and offer more support than client-based applications. In most cases, both web-based applications and commercial websites have a database such as Oracle, SQL Server, MySQL, or Access (in rare cases) as the back end. Fortunately, no matter what the type of database, the same process and tools (e.g., Multilizer) are used for localization.

From an engineering perspective, the most important step in localizing a database is to use well-defined spec sheets that list the tables and the fields requiring localization. It also helps to have the database designed in such a way as to facilitate either field or table localization. From there, the only other hurdle could be string length limitations, but these are easily managed with tools such as Multilizer.

As with many things in life, however, what is good for the goose may not be good for the gander. Websites built with dynamic content are usually very localization-friendly for engineers; in most cases, it is relatively easy for us to extract the text strings from the underlying database. Unfortunately, once the text has been extracted, it is not so friendly for the linguists who translate the strings.

Mike van Grunsven
Senior Localization Engineer

Never in my wildest
dreams growing up did
I imagine that I'd be a
Localization Engineer,
let alone a senior one.
Seriously though, who even
thinks about that kind of
thing when they are a kid?
But I'm glad for my job. The
variety of tasks that comes
with it really keeps you on
your toes. Here at Lingo
Systems the engineers have
their hands in every job
that comes through our
doors, so we get to see
a little of everything. The
diversity of products and
technologies is astounding.
I can't think of a day that
goes by that I don't learn
something new. So, is it
a dream job? Not in a
traditional sense. Would I
give it up? Not a chance!

Rather than working with a complete document, all the translator sees are random, out-of-context strings—a difficult challenge for even the most skilled professional. It is therefore a good idea to use a description field in your database to give some guidance to the linguists. With proper instructions, the engineer will be able to include non-translatable fields in the translation packages that are provided to the translator.

The most compelling advantage of a database-backed website is the downstream benefit. Updates (including localization maintenance) become very easy and very cheap. As changes are made to the site, the new and modified strings are extracted, translated, and then reinserted. In many cases, localization delivery can even be automated using a translation management system such as Lingo Systems' Lingo*NET*. See Chapter 19 for a complete description of translation management systems.

There can be other challenges to localizing a website besides the database component. For example, using multiple programming languages can create parsing difficulty when generating RTF files for the linguists. The most common programming languages found on the majority of websites are PHP, JSP, Perl, ASP, ColdFusion, and all flavors of Java (the programming language, not the coffee).

Last, but far from least, working on the graphical assets of a website can be difficult when source materials (Photoshop, Illustrator, or CorelDraw files) are not available. With their omnipresent gradient backgrounds and obscure fonts, nothing is worse than being asked to recreate localized versions of these elements. This invariably requires design expertise from (and budget for) the DTP department.

Repeat after us: It is always a good idea to keep the source files in a safe place and to isolate localizable layers. This useful feature is offered by most, if not all, image and graphic editing software such as Photoshop, Fireworks, and Illustrator. Another smart approach to graphics is to use PNG as a source format as this format is text-based and allows the localization vendor to translate directly inside of the graphic without the need for a graphic editing application. The only limitation of the PNG format is that it cuts multi-line sentences into several strings.

Talking the Talk: Terminology Management and Translation Memory

The last function that a localization engineer performs may be the most important. Terminology management, including the creation and maintenance of a translation memory (TM), has a huge effect on both quality and consistency. It may also be the single most important factor in reducing localization costs.

Translation memories are a must-have for any localization project. Some localization firms assign the task of TM and terminology control to the project manager. At Lingo Systems, we believe in using the right person for the job and have no doubt that when it comes to managing hundreds of multilingual translation memories this person is the localization engineer. An inaccurate or corrupted TM (whether it is a linguistic corruption or an encoding corruption) can reduce leveraging, adding to the cost of the project and ultimately hurting the quality of the translation.

There are several players in the CAT tool market. Since the acquisition of Trados by SDL in June 2005, the largest share belongs to SDL TRADOS. But this acquisition opened the doors to numerous new players such as Heartsome, Swordfish, Across, and DVX (version 10 of DéjàVu), just to name a few. The principle behind each of these products is the same: The translator uses the tool interactively within a word processor to automatically retrieve existing translations from a TM. For localization engineers, it does not really matter which tool is used since most, if not all, are TMX compliant, meaning that the TM content can be exchanged between CAT tools through an XML-based export file (TMX file). All of them also offer fuzzy matching, which gives the translator close matches to a localizable sentence, thereby speeding up the translation.

Another linguistic tool that is often integral to the development and maintenance of an effective TM is glossary management. In the world of localization, glossaries represent a list of key terms and definitions that the translator will need to properly localize the source materials. Many of the TM tools include a glossary management module to facilitate the compilation of a glossary and the subsequent translation of the key terms whenever (and wherever) they appear. These modules, such as MultiTerm from SDL TRADOS, run in the background as the translation is being done in a word processing application. They then flag any term that is located in the MultiTerm glossary, minimizing the time a linguist needs to go back and forth between reference materials and applications.

The newest glossary management tools are so customizable that they even allow the user to add multimedia content to the term definitions. The possibilities are infinite. The next generation of translation tools even allow the localization vendors to share their glossaries as well as their translation memories over the Web (SDL Trados TM Server and MultiTerm Online are good examples), which greatly facilitates the interaction between the localization company, the client, the linguists, and the in-country reviewer.

Moving Forward

Through countless acquisitions and an ever-increasing need for translation, the localization industry is moving at a very fast pace. The best partner to go global with is one that is active in shaping the future of this industry and whose staff is experienced and knowledgeable in the use of the latest tools and technology, can master new formats, and is involved in beta testing programs.

The direction in which our industry is going is clear and logical. As globalization is expanding, every company that is serious about its financial future wants to get a piece of the global marketplace. Even though this seems like an expensive proposition, it does not have to be. Your localization partner should understand that every penny saved on a project is a penny that you can spend on extending your global reach. The cheaper it is for you to localize into one language, the more languages you can localize into. It is a win-win situation, both for the customer who gets access to more locales and increases potential revenues, and for the localization firm that increases its revenues by localizing more languages.

The first step in reducing localization cost is to control the source content. A CMS will allow you to increase productivity and reduce costs in many different ways:

- Through the use of an authoring memory that presents your authors with sentences that they have already created and tells them in which languages they have already been translated. This is the same concept as predictive text on a cellular phone.

- By allowing only new and modified text to be provided to the translation vendor. It is important that the linguists get contextual information to mitigate mistranslations.

- By eliminating the need for DTP through the separation of the text from its formatting.

- By offering all necessary outputs through the use of customizable templates.

If the CMS you are evaluating does not have the above features, you probably should look for another one.

Once you have controlled your source content, you have to get it translated. When doing so, it is important not to just look at the price you are going to pay for each word, but rather the value that price represents. For example, why pay for a translation that might have already been created by one of your competitors? One theory floating out there is to allow a customer to pay less per word with the agreement that the resulting translation be included in a Global TM (GTM) that can be used by other customers to increase leveraging. Only customers who agree to participate in the program

would be allowed to benefit from the GTM, which could very well be industry-specific to reduce mistranslations. The same approach can be taken at the glossary level. This step would be taken by the localization company and would have no impact on the translation teams except for a reduction in the number of words to be translated for a given project. It's important to note that many companies who cannot currently afford to translate their materials but might be able to do so through the use of this controlled crowdsourcing method.

Another area where productivity can be increased is in how the actual translation step itself is done. Most technical translators have already used machine translation for several years in order to increase their throughput. A single skilled translator can increase translation volume up to 7,000 words a day (this is not an estimate but actual data)—even on the low side, an increase to 4,000 words is a huge leap compared to the current 2,200 words daily standard. So, it is no surprise to those in the know that the latest version of the Workbench now offers a machine translation module for all unknown segments. If only 15% of those segments are perfect matches, it still adds up to big savings. As a translation community, we have to embrace machine translation and expand its use to actual production work. See Chapter 20 for a comprehensive discussion of machine translation.

Wrapping Up

Let's speak plainly. Localization engineering is not rocket science, but in our estimation, it comes close. As you prepare for a localization project, be sure to leave a seat at the table for an engineer. From the initial internationalization planning, through actual translation and implementation stages and on-going translation memory development and maintenance, an engineer should be directly involved. We may be biased, but we believe that a top-notch engineering department can anticipate the potential issues you may face well ahead of time, from nagging technical oddities to esoteric cultural differences. So, as you plan your next project, keep us engineers in mind. The extra time you invest up front will pay off in terms of reduced timelines and cost savings in the long run.

oops

A menu from Bolivia:
Weed bread with fried cheese

Dawni Jacobs
Business Development
Manager

Even before I was working with creative agencies I was using the term "flash," albeit in a different context. Raising five children meant everything had a flash component. However, in the multimedia arena the word flash doesn't exactly mean "quickly" and "right now." Since its introduction in 1996, Flash has become a popular method for adding animation and interactivity to web pages; Flash is commonly used to create animation, advertisements, and various web page components, to integrate video into web pages, and more recently, to develop rich Internet applications. My former years at creative agencies have proven to be an asset here at Lingo Systems when localizing rich content, animation and videos.

Chapter 10

Multimedia Localization

As communication tools have become more sophisticated, the methods used to localize their output have had to change as well. Not too long ago, localization was most often performed on written materials (e.g. product manuals, user guides, training materials, etc.). As the world became wired, next came online help, websites, and GUI. The new frontier, which is already everywhere, is multimedia. Some of the more common examples of multimedia materials are Flash movies, sound files, video clips, and complex graphics. All can be found in such diverse places as video games, interactive software, web applications, DVDs, kiosks, and CD-ROMs.

Unlike traditional localization, which typically uses professional linguists to translate the materials and desktop publishers to format the output, multimedia localization requires experts with additional skill sets for audio or video text adaptation, script translation, and professional voice talent recording. A healthy dose of software engineering is also needed. This combination of linguistic work and in-studio production services can be used for training, marketing, educational or commercial audio and video applications, as well as entertainment products.

Regardless of the medium, there are a few basic rules that should be followed on every project. In this chapter, we will provide a brief overview of multimedia localization and describe a sample Flash project.

Pick Your Format

When it comes to presenting your multimedia message, the choices of format are seemingly endless. Your production team can choose from a wide variety of applications that use many different formats, including:

- 3D Studio Max,
- Alias,
- Animated GIFs,
- DirectX,
- Macromedia Flash,
- Macromedia Shockwave, and
- SoftImage.

Once you've decided on your format, the next decision is whether to provide your multimedia presentation to a global audience. If your message is intended for multilingual consumers, make every attempt to consider localization from the outset, when the source files are first being developed.

Begin With Localization In Mind

Communicating to a global audience with a multimedia project presents both technical and cultural challenges. If you begin with localization in mind, you can avoid frustrating delays and increased costs down the road.

On the technical side, you may run into issues such as text expansion. Plan for a minimum of 25%, which applies to on-screen as well as spoken text. Make sure you understand the character sets for different languages and how issues such as concatenation and bidirectional scripts (see Chapter 15) are handled. Anticipate how the synchronization of voice files will differ between languages. Can your on-screen buttons accommodate translated text? Make a final linguistic check on all these items.

As with written words, visual messages may inspire very different meanings as they cross from culture to culture. For example, an image that is acceptable in Europe may be unacceptable in Asia. Use care here: A misstep might result in some embarrassment, but it could also detract significantly from your message and negatively impact your bottom line.

How Localization Gets Done

Once you have created your multimedia message in its source language, localization begins with a localization engineer (carefully) dismantling your lovingly constructed source document, keeping an eye on such issues as color specifications, system fonts, text attributes, navigation and interactive text. On-screen text is then pulled out and sent, along with any scripts, for localization.

After translation, it is of paramount importance to have the translated on-screen and narrative text approved prior to entering the studio or starting engineering. At the very least, this helps avoid basic cultural and industry issues. At the most, it gives your in-country team a valuable stake in your localization project. Additionally, their up-front buy-in avoids downstream delays to the timeline and costly studio and engineering re-work.

Once approved, the localization engineer places the on-screen text into the localized visual files. Then the script goes to an audio/visual team that may include audio engineers, voice talent and, perhaps, a linguistic producer. Working with the specifics of the source files, this team produces the localized voice files that are handed off to the localization engineer for final placement.

A final engineering step to spot script errors and broken links plus one last linguistic review to ensure synchronization, and your files are ready for a global launch. By keeping these steps in mind from the outset of your production cycle, you will guarantee your multimedia project entertains and enlightens every member of its audience.

Flash Case Study

The Product
Interactive computer-based training modules

The Client
One of the nation's largest retailers

The Project
Localization of computer-based training modules

The Scope
- 16 Adobe Flash files (84 fields to localize)
- 7 short WMV files (subtitles provided/total length: 5m14s)
- 6 graphics to localize (less than 200 words)
- 13 MP3 audio files to localize
- 4 MS Word documents (scripts for audio narration and HTML on-screen text)

The Challenges
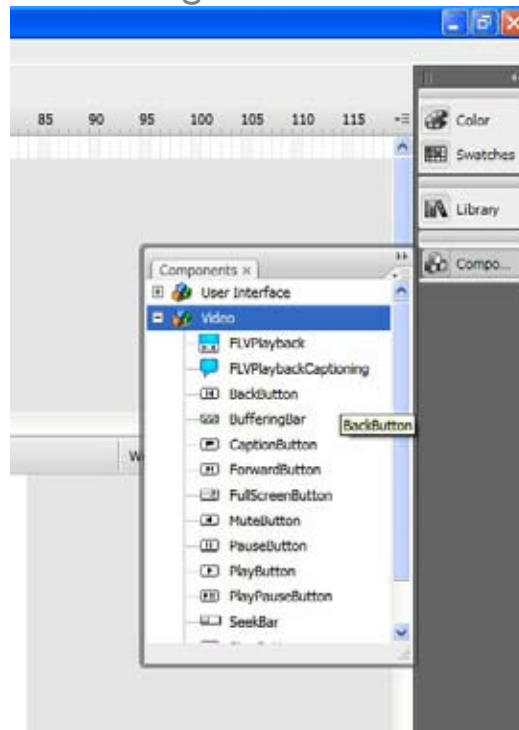
During the estimation process, localization engineers determined that the subtitling for the video should done via the Adobe Flash subtitling engine using external XML files for synchronization and implementation of the text content. This solution used the built-in flash components "FLVPlayback" and "FLVPlaybackCaptioning" on the Adobe Flash CS3 platform.

The second engineering challenge for this project was that the course scripts did not have a naming convention for the audio files and some of the text content for the Flash material was missing.

The source files did not contain the script for the seven training video files. In order to localize and provide the subtitling for these movies, we needed to extract a transcription from the English source file and translate it, then use that translation as the content for the subtitles of the seven training videos.

The Process

In order to localize and provide the subtitling for the videos, we extracted the English source text and translated it to implement the subtitles. The translated content was transferred into a Timed Text (TT) XML document which looked like this:

```
<style id="default" tts:backgroundColor="black" tts:textAlign="left"
tts:color="yellow" tts:fontSize="16"/>
<style id="associate" tts:backgroundColor="black" tts:textAlign="left"
tts:color="white" tts:fontSize="16"/>
<style id="customer" tts:backgroundColor="black" tts:textAlign="left"
tts:color="yellow" tts:fontSize="16"/>
  </styling>
 </head>
 <body>
  <div xml:lang="es">
   <p begin="00:00:01.00" style="associate">
Hola.
   </p>
   <p begin="00:00:02.00" style="customer">
Hola.
   </p>
   <p begin="00:00:03.00" style="associate">
¿Cómo va a pagar esto?
   </p>
   <p begin="00:00:05.00" style="customer">
Voy a usar mi Acme's charge.
   </p>
   <p begin="00:00:11.00" dur="00:00:04.00" style="associate">
¡Ah, claro! (por el altoparlante) Se necesitan más cajeros al frente.
¿Marqué esto?
   </p>
   <p begin="00:00:21.00" style="customer">
No, usted marcó esto. Ahora, esto está en oferta, ¿correcto? Sin
embargo, tal vez no lo quiero.
   </p>
```

The trained eye will notice that different styles have been created for each character in the movie; the subtitle for the associate displays in white and the customer displays in yellow. This is just an example of how subtitle styles can be used to improve readability. There was also the possibility to set up a duration time for each subtitle, if desired, to avoid seeing the subtitle on the screen for too long.

A linguist gave the exact timing of the subtitling, and with the help of the localization engineer, that timing was implemented in the Timed Text (TT) XML.

The Solution

Because the source files for the movie were Windows Media files (.wmv) but the shell of the course (CBT) was done with Flash, the decision was made to use the "FLVPlaybackCaptioning" component that came by default with the latest version of Adobe Flash (CS3). Because the subtitling data was sitting in an XML file, it was very easy to update the files (versus hard coded or even embedded subtitles). Most important, it was quite simple to translate the XML file into another language, especially if we wanted to localize the CBT into many other languages.

The "FLVPlaybackCaptioning" component relies on the "FLVPlayback" component to play the movie itself, but that component can only handle FLV files (Flash video). So in order to use that method we converted the WMV files to FLV.

In conclusion, the use of the Flash component for subtitling purposes allowed flexibility and was easy to implement with few or minor drawbacks.

PROS:
- Easy to implement
- The use of XML made it easy to localize
- Easy to amend

CONS:
- Worked only on FLV movie format
- Required the latest version of Adobe Flash (CS3)
- The client needed to support the use of Flash technology
- There may have been web server parameters to implement if the movie was to be viewed online

For the rest of the localization, Flash Shell and miscellaneous graphics did not present any particular challenge. The Flash Shell used a mix of hard-coded text fields and dynamic fields linked to an external XML, and the localization of the graphics was done in Photoshop.

The key decision for this project was the selection of the proper technology to process the subtitling and the synchronization of those subtitles. Despite the challenges, this retailer had the right partner and the job was completed on time and on budget. If you follow the lessons learned in this chapter, getting to the desired outcome will be a breeze for your team.

Choose Your Type

There are two choices for how to use text when creating a source Flash document. Here are some things to think about when it comes to localization:
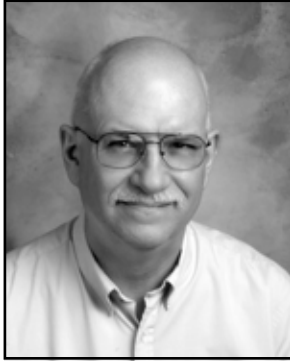
Regular Text field: This is the easiest to manipulate and localize. Localization software can automatically find this text, pull it out and leverage it against any existing translation memory. The localization engineers still need to gather information about the fonts and attributes used in creating the source file, *and* implement the fonts into their system prior to publication. Otherwise, Flash will automatically substitute for missing fonts and undesirable results may be published.

Bitmap Text field: While these text fields cross platforms and versions with ease, they can be a challenge for the localization engineering team. The resulting text is not editable and the only solution is to manually enter each localized string into the same authoring tool using the same settings as were used in the creation of the source. Bitmap text is often found when the source designer has used a third-party authoring tool to apply effects to type.

Notes: Text Field Type Pros & Cons

	Pros	Cons
Regular Text field	Editable Can be linked to an external XML file	Portability issues can occur
Bitmap Text field	What you see is what you get (WYSIWYG) 100% portable	Not editable Necessary to have the source font

Chapter 11

Testing and Integration of Software

Software localization is a complex process. Any number of issues can arise—grammatical, technical, or cultural—when translating and deploying a software application across different languages.

Testing the software after localization is an essential quality assurance procedure that is designed to discover, document, and communicate issues that could compromise an application's integrity when it has been translated for a local market. Through rigorous and thorough testing and reporting, your vendor's testing staff should attempt to reduce damaging uncertainty about the perceived state of localized software. This goal can be accomplished through the "4 Rs": Review, Reveal, Record and Report.

Review

In the *review* phase, your vendor should work to help you understand the nature of the software or product under consideration for testing. Among the factors that that should be considered:

- The extent and variety of localization and internationalization desired,
- Languages, regions, and character sets that are required for proper localization,
- Interfaces and files associated with the application, and
- The width and depth of testing coverage desired, in accordance with your time and budget constraints.

When all of these issues have been addressed to your satisfaction, the test plan is started. It will evolve in scope and complexity as each of the following phases is considered and completed.

Reveal

The *reveal* phase looks at and exposes the software or product behavior. Within it are the "4 Ts": Type, Technique, Test, and Target.

Type:

- **The pseudolocalization type** is best used early in the development phase to reveal truncated characters, clipping, and line wrapping issues before significant efforts have been invested in formal localization. Hours spent here can save weeks later. By inserting and modifying characters within the software containing boxes, various user interface (UI) defects can be exposed—such as unforeseen space considerations when a translated word or phrase is longer than its source form.

Dick Pilz
Testing Coordinator

Originally from Chicagoland, I've lived most of my life in "Munich on the Willamette" or Beervana. I know where some of the best beer and mead in the world is made, which includes some of my own. I love tinkering and solving puzzles, whether hardware, software or recreational. At Lingo, I ride herd on a collection of computers with multiple personalities together with our test technicians. Our job is discovering, documenting and communicating to our clients through tests and reports to reduce damaging uncertainty about the perceived state of localized software.

Think of it as if you are building a house. Spending extra time up front planning the dimensions of your doorframes and living spaces can spare you untold time, inconvenience, and expense, rather than discovering that, after the walls and doors are hung, you cannot accommodate a sofa.

- **The localization type** looks at the consistency and appropriateness of the software for its intended users and market. It typically is used on the release candidate of the software. For example:

  - Units of measurement: Are the units of measurement presented in the application appropriate for the target market?

  - Date and time: Are the date and time format locally correct?

  - Non-localized items: Are there any trace elements within the application that have not been localized?

  - Fonts: Have default fonts translated properly into the localized version of the software? Do all characters present clearly and appropriately?

  - Menus: Is there menu consistency between a localized software version and its source version? It's not uncommon for a different number of menu items to present between various localized and source iterations of an application.

  - Extraneous content: Does extraneous content appear in the localized version of the software that did not exist in the source?

- **The linguistic review type** focuses on language correctness. It can occur early in the development phase for the bulk of the review and again in the context of the functional UI. Your localization vendor should employ highly qualified and experienced professional linguists, fluent in both the source and target languages, to thoroughly review localized applications for accurate and appropriate translation. They should be able to assess:

  - Missing content: Did all content from the source version transfer to the localized version? Omitting key words or sentences can severely impact a localized application grammatically and functionally.

  - Spelling errors: Are all of the translated words spelled correctly in the localized software?

  - Accuracy of translation: Is each word or language segment viewed and edited according to the highest standards of the target language? Slang terms and lower idioms should not be included in the localization.

  - Functional translation: Because grammatical structures vary by language, literal and word-for-word translations are inappropriate for localization efforts, and often only confuse the native speaker of the target language.

  - Translation consistency: Is a product or company name consistent throughout the localized version? Do frequently used functions appear consistently throughout (words such as "Close," "Next," "Edit," "File," etc.)?

Neil Kerr
Localization Engineer

What happens when you get your degree in Anthropology and Linguistics; spend 10 years of your adult life as an expatriate in Prague; discover an interest in "informatics"; and use words like "decontextualize" and "bricoleur" in everyday conversation? You become a localization engineer.

- Language style: Have technical processes and common terms been translated in a way that is consistent with the source's language style? For example, where you would use "sheet," "page," and "paper" interchangeably in English, in another language "sheet" and "paper" may convey different meanings, and only "paper" would be appropriate.

- Imperative, or command, voice: A command such as "Click to Proceed" should be translated in the same voice as other commands (and not as "You should click the button to proceed to the next page").

- Punctuation: Punctuation should be translated according to the punctuation rules appropriate to the target language.

- Local convention: Sensitivity to local convention is very important in any localization effort. Some phrases, words, and gestures that are considered harmless domestically can be considered offensive in some target markets.

- **The internationalization type** focuses on the function and appropriateness of the software when exercised on localized and regionalized systems, regardless of whether the software is localized. For example, an English product should install and function on a Japanese system. It should also recognize and display Kanji characters and Imperial dates.

- **The functional type** is often combined with the others in this phase. Installation, work-flow and error handling are explored here. In past testing efforts, this step has exposed error conditions unique to a particular localization of an application—a printing error that occurs only in a Spanish version of an application, for example, or an installation error message that persists in all non-English versions. Operations most frequently considered in functional testing include:

  - Install, installation modification, and uninstall in different configurations (for example, a US English installation on a Dutch-localized machine).

  - Menu functions and Hot Keys.

  - Help links and file paths.

  - Core application functions.

  - Localized input device functionality (international keyboards, for example).

  - Regional Default Settings: Do settings for items such as paper size, date/time, punctuation marks, and currency formats default to the appropriate settings for the country or region for which your machine is localized?

  - Cutting, copying, and pasting special characters: Do they translate correctly on a localized system?

  - Document transferability: Can saved documents created in the source version be used in the localized version?

- **The regression type** focuses on verification of defect fixes and modifications. Has a previously-reported error condition been corrected in the latest iteration of the software?
- **The Section 508 type** is focused on usability for people who are impaired for display or input, such as alternative keyboard navigation. It tests compatibility with assistive technology, such as magnification, screen readers, refreshable Braille displays, and speech recognition tools.
- **The web accessibility type** is related to Section 508, adding in browser compatibility and function.

Technique:

- **White box testing** uses an internal perspective of the software to design test cases based on internal structure. It requires programming skills to identify all paths through the software. It is extremely useful for testing localized web applications. The UI is not involved, as in some linguistic reviews.
- **Black box testing** takes an external perspective of the test object to derive test cases. These tests can be functional or non-functional, though usually functional. Valid and invalid actions are selected, based on implicit or explicit requirements and determines the correct output. There is no knowledge of the test object's internal structure.
- **Grey box testing** is a hybrid of the two techniques above where the internal data structure and algorithms are known, but the testing is done at the UI or user level.

Test:

- **Automated testing** is best used when many tests are executed repeatedly, but the software does not change much between releases. The test is often created using a script by recording and capturing inputs for later playback and review. As an example, a series of actions and inputs may be executed on an English system. Later, the recorded actions and inputs are played back on a localized system. This is very useful for conducting linguistic review in UI context. It can also reduce tester error since it restricts input variability, but that can be a mixed blessing in a localized context.
- **Manual testing** is very flexible, but it is best conducted with experienced testers. There are several types of manual tests.
    - Scripted tests have a fixed set of steps, without necessarily requiring a typical usage.
    - Scenario tests model user sessions or practices, such as opening, modifying, saving and exiting files.
    - Ad Hoc tests are free form and unscripted, driven by tester experience, but often with a goal in mind.
    - Exploratory tests roam the area between scripted and ad hoc tests.
    - Smoke tests are brief, touching only the most common "happy path" areas. This is most often conducted immediately after a build or release of software.

Target:

- The target is the environment in which the software is intended to be used. It may only consist of a computer, but can also include other devices.
    - An **exact machine image** of an operating system version at a specific patch level is often required for verified, audited testing.
    - A **virtualized machine image** is often required for white box testing development, especially web applications.

These images may be localized with an operating system specific to a language and may be further modified for a specific region, such as Mexican Spanish or Brazilian Portuguese.

Record

A *record* of the software behavior and UI is necessary evidence for reporting and for future reference. All records should have provisions for signing and dating.

- **Video capture** produces a movie of the inputs and UI displays.
- **Screenshots** of specific UI may emphasize features for later communication.
- **Electronic or hardcopy logs** may be kept. Often, manual scripts have entries for logging.
- **Print captures and print hardcopy** may be produced, especially for recording report outputs.
- **Documentary evidence** of all error conditions (error messages, screen captures, error logs, etc.) and text execution are essential for accurate and detailed reporting of potential defects and functionality issues.

Report

A *report* is the deliverable of localized software testing. It has both interim and final forms.

Interim Reports:

- **A defect tracking report** is one type of interim report. Records and scripts are entered into a database for review and action by the customer.
- **A progress report** is another type of interim report with metrics for gauging time and resources expended towards a goal.

Final Reports:

- **A validation report** documents evidence that provides a high degree of assurance that a product, service, or system accomplishes its intended requirements.
- **A verification report** documents evidence that provides a high degree of assurance that a product complies with regulations, specifications, or conditions.

Summary

With this coordinated and methodical testing approach, your vendor will ensure that a localized application can be released seamlessly into a target market. At the core of the testing process, your vendor should attempt to reproduce the real-world experience of a local user as closely as possible. While they may not be sitting down at a personal computer in an Istanbul home or a laptop in an Internet café in Lisbon, your vendor should attempt to anticipate every behavior that such a user could encounter when the localized application is introduced to their local environment.

OOPS

Signs in Beijing:

Offer the seats to the old, weak, sick, crippie and gravid

Those who suffer from high blood pressure, mental disease, horrifying of highness and liquor heads are refused.

Very Suspicious Supermarket

Lettuce in Pain

Please don't touch yourself. Let us help you to tryout. Thanks!

Roger Thompson
DTP Supervisor

I have a friend, who's got
a friend, who's seen a
Sasquatch in the Cascade
Mountain range. Now you
know it's real.

Chapter 12
Formatting Print and Online Documents

Desktop publishing (DTP) is an important part of many localization projects. It is not enough to simply translate the words and let them fall where they may. Care must be taken to present your manuals and help systems with the same sense of style and polish that they have in their source language. Fortunately, the latest technology makes this task much easier. In this chapter, we will review both printed and online documentation, the changing nature of fonts, the wiles of text expansion, resizing screen captures, and portable documents. We end with a look at the tools that have revolutionized "tech pubs" creation: single-source and content management.

Experience Counts

Localization vendors handle desktop publishing in one of three ways. Some use their own employees, others outsource to contractors (either off-shore or on-shore), and the rest subcontract with the translation agencies who perform the linguistic work.

The differences can be significant, so do your homework. A skilled formatter knows many ways to ensure that your materials look as good in the new language as they did in the original. Also, with expertise comes efficiency. While the neophyte is searching for some way to squeeze text elegantly, the old pro simply reaches into a bag of trusted tricks. You shouldn't have to pay for someone else's on-the-job training.

Whether your vendor's DTP resources are located off-shore or on-shore can also affect your project. When timelines are critical or project complexity requires close coordination with other departments such as localization engineering or QA testing, there are often significant advantages to keeping these tasks in-house.

The best way to ensure that you receive well-formatted deliverables is to ask for references and check them out. Inquire as to how many years of experience the formatters have with the desktop publishing program you use. If the job requires a new program or a new version of a traditional program, make sure the vendor knows how to use it and that they have the appropriate target language version. Since the formatting for some languages can only be done on a native operating system, your vendor will need to have invested in a variety of systems and licensed publishing tools. Even experienced formatters can get lost when working with other

languages if they are only used to their own native tongue. On the other hand, experienced localization vendors will already have dealt with most system, application, and file compatibility issues. If you are invited to visit the vendor's site, ask for a tour and a demonstration of capabilities.

Selecting the Right Font

Software fonts have been around for more than twenty years. The earliest versions were sufficient to display text in dot patterns on monitors and dot matrix printers. Most did not have an extended character set to display accented characters, which is a fundamental requirement for localized documents. Even today, some of the new specialty fonts do not feature even the simplest accented vowels. If your document uses one of these fonts, the font will probably have to be replaced to display in other languages.

Multilingual fonts were first included with the Windows operating system in the mid-90s. If the target languages were European, Slavic, Baltic, Cyrillic, or Turkic, the available fonts in the Windows character map application would probably suffice. Double-byte languages were another matter. Hebrew, Arabic, Hindi, and most Asian languages required font substitution even when Windows multilingual fonts were used.

The twenty-first century has seen a breakthrough in software fonts with the inclusion of Unicode in the Windows XP and Macintosh OS X operating systems. Because Unicode fonts use more bytes per character, most alphabets around the world can be represented, save for extremely obscure alphabets such as those using cuneiform script. Even if an operating system supports Unicode fonts, there are some legacy DTP applications that do not accommodate them, like MacWrite in pre-OS7 Apple computers, or MS Word in DOS 3.1. Soon all software fonts will be Unicode and applications that do not display them will be defunct, if they are not already.

Fonts can also be an issue when making PDF (portable document format) files. In the early 90s, some font manufacturers took pains to prevent their fonts from being embedded out of the fear that hackers would be able to extract them from the PDFs. This concern never materialized because the effort did not equal the prize. But if a document uses one of these older fonts and you wish to make it portable, consider changing to a different font.

Text Expansion

When English is translated into other languages, it often takes more space to say the same thing. The reason may be that the new language uses more articles, as in French or Italian, or because the words are simply longer, as in Dutch or German. On the other hand, the new language might use a few ideograms to express an entire phrase causing the opposite effect. While text contraction is rarely a problem when localizing documents, text expansion

TEXT EXPANSION/ CONTRACTION	
Language	% difference
Arabic	104%
Chinese	61%
Czech	117%
Dutch	128%
English	100%
Esperanto	92%
Farsi	104%
Finnish	103%
French	111%
German	108%
Greek	128%
Hebrew	83%
Hindi	83%
Hungarian	113%
Italian	109%
Japanese	115%
Korean	123%
Portuguese	110%
Russian	115%
Spanish	117%
Swahili	88%
Swedish	95%

(George Sadek & Maxim Zhukov, *Typographia Polyglotta*, New York: ATypI / Cooper Union, 1997. The study compared the Preamble from the Universal Declaration of Human Rights in a variety of languages, with English as the base 100%.)

Dianne Ellis
DTP Specialist

My family and I are totally dog people. I think we believe we can really communicate with our furry friends. But if our dogs were in another country they would sound completely different. For instance, in Russia they would be saying "How-how" rather than "Bow-wow." Or "Wang-wang," "Mong-mong," "Huf-huf," "Ar-ar," or "Wah-wah," depending on the country they are in. I just know that if you get 6 of them under the dining room table at the same time it is very crowded.

can raise some tricky issues. The standard rule of thumb in the localization industry is that European languages expand, on average, by about 30% (without hyphenation). This can cause several challenges. A table that fits on one page in English may spill over to the top of the next page in Greek. Similarly, section headings in large type might run to two lines. Indented text could leave large blocks of white space to the left.

At the beginning of the localization project, your provider should ask questions to determine how you want to handle text expansion. Do you want to shrink the font to ensure that pagination, TOC entries, and index references match the English source document? If so, should the line spacing (or leading) shrink proportionally? Or would you prefer to keep the same font sizes and allow the text to flow, increasing the page count? Can formatters "borrow" space from the margins? Can the indentations be shrunk to reduce white space to the left? Can headings be made smaller? If so, should it affect all similar headings or only those that present a problem? When long words expand to accomplish full justification, the spacing between the letters can stretch in ways that are uncomfortable to read. Can justification be turned off? Do you prefer to avoid hyphenation? In some languages this results in sudden line breaks. By addressing these layout concerns at the beginning of the project, your localized documents will have much higher quality when they are delivered. See Chapter 13 for more information about page design and text expansion.

The table on the previous page presents an interesting look at one study on text expansion. While the information is informative, the percentages can change significantly depending on how you decide to handle compound words and hyphenation.

Online Documentation

Online documentation avoids some of the pitfalls of text expansion and page matching that are associated with printed materials, but it introduces engineering issues. For example, if your document is displayed on a computer screen in HTML, WinHELP, or some other online format, expansion will not be a problem since the text will extend downward and the user simply scrolls down the page to read the "expanded" text. What can be an issue is whether or not the content displays correctly on the operating systems and browsers available in your target market. By performing functional testing on native operating systems, your localization provider will be able to ensure that the applications perform and display as advertised. Be sure to discuss your specific engineering testing requirements with your localization vendor so that you are both clear on testing expectations.

Resizing Graphics and Forms

Online forms and other page elements that contain text such as graphics and buttons may also need to be resized after translation. Similarly, online forms may require special engineering to support the user's ability to enter special characters, international phone numbers, and foreign addresses (along with any other special requirements requested by your international users).

Screen Captures

Almost all software documentation uses screen captures, which are no more than pictures or graphics of the software as displayed on screen. This has become a technical documentation convention for tying together the references in the document to what the user sees on screen.

Just as translated text expands in the body of the document, translated dialog boxes in applications may expand as well. One of the most common examples is an error message box. A screen capture of an error message that was originally 432 pixels wide (6 inches on a 72 dpi Mac system), might need to be expanded to 504 pixels (7 inches) to fit the translated text. In the documentation, the 7-inch screen capture must be resized to fit in the same space as the original 6-inch screen capture. This often leads to distorted or fuzzy images. To solve this problem, the screen capture can be left at 7 inches (possibly throwing off the formatting), resized to something that causes less distortion, or used at 6 inches, accepting the distortions. If your document uses screen captures, be prepared to talk with your localization vendor about how you would like these matters resolved.

It is also important that the screen captures of your localized software be taken on a localized operating system. If a question is asked in French, the "yes" and "no" buttons should be French as well. For more information on localizing screen captures, see Chapter 9.

OOPS

Advertisement of a Hong Kong dentist:

Teeth extracted by the latest methodists

Diana Warnock
DTP Specialist

Language shapes perception.

Perception shapes reality.

Reality is social construction, primarily built with language.

Humans are the only species having the capacities to communicate thus.

As a desktop publisher, I have opportunities to see our differences and commonalities; I believe language is the bridge across economic & cultural disparities to global nonviolence & peace.

Portable Document Format (PDF)

With the advent of PDFs, documents became fully "portable" with their original layout intact. Over time, PDFs have become specialized according to how they will be viewed. A high-resolution PDF is necessary for high-quality offset printing, while low-resolution is best for the Internet. A print-quality PDF is usually huge in file size, often over 100 MB, while an online PDF can be as small as 1 MB or less. In both cases, the fonts must be embedded. Another version is a functional PDF, which contains bookmarks and links, and can even launch other applications. It is important to indicate which kind of PDF you want at the beginning of your project.

Single-Source Content Management

The concept of single sourcing predates true content management. As the words suggest, it means producing multiple outputs from one source document. With the use of hidden text, variables, conditional text, and export filters, a desktop publisher is able to publish to multiple media, including hard copy documents, web pages, electronic books, help files, and more.

Single sourcing can be as simple as using the "Save As ..." command from MS Word to create web pages. In this case the source is the Word file and the hard copy and web page are two different outputs from the same (single) source. Anyone with a copy of MS Word on their computer can accomplish this. If the web pages are subsequently imported into a help system, you now have a third use of the same material. When localized, only the Word file is translated. The original file then becomes a translation memory from which many other deliverables can be produced.

If your publishing requirements involve anything beyond limited Word documents, or if localization is in your future, the quick fix described above will not be sufficient. Thankfully, far more powerful and sophisticated single-source and content management solutions are currently available. These range from off-the-shelf single-source publishing programs to highly customizable, enterprise-level content management systems.

For single-source publishing, Quadralay Corporation's suite of WebWorks products enables users to convert output from MS Word or FrameMaker into HTML, WinHelp, HTML Help, or even XML. The real tricks are in what can be done with the text as it passes through WebWorks. For example, a skilled user can map the styles in your source to any other presentation style by using a cascading style sheet (CSS). Or, you can create character maps to match special and extended characters to their corresponding HTML codes. As with any application, the more features it has, the more complicated it is to use. Fortunately, WebWorks comes with some templates that can be used for simple operations. Adobe's RoboHelp product has similar output capabilities and has traditionally been popular with help authors.

A single-source solution that also includes authoring and content management tools may be appropriate if, in addition to multiple outputs, your publishing environment includes any of the following:

- multiple authors,
- a desire to publish in many languages,
- a high degree of similarity between content, and
- difficulty in managing version control.

Not too long ago, content management systems were practically unobtainable for all but the largest corporations. Now, state-of-the-art applications such as Author-it (which was used to create and publish this book), Vasont, and Interwoven (to name a few) use object-oriented, or relational database architecture, to combine multimedia publishing capability with authoring, version control, file sharing, and a host of other features. Perhaps the best news is that there are content management systems to fit virtually any budget, from highly customizable enterprise applications to entry-level systems with full functionality.

The need to provide clients with localized content is a major reason why so many companies are implementing content management systems. These tools provide a framework for creating and maintaining control of multi-language content, thus minimizing cost. A more thorough discussion of single-source content management tools is presented in Chapter 17 by Paul Trotter, CEO of Author-it Software Corporation. This is followed by a description on the many benefits of integrating CMS tools into your localization workflow in Chapter 18.

oops

Misquoted in the news:

In February 2009 the Hungarian Prime Minister said in a speech, "The world is in deep trouble and Europe too, but there is a Hungarian phrase that you get to know who your friend is when you are in trouble." The newswires quoted him with the mistranslation "we are in deep trouble, indeed," which caused the Hungarian *forint* to decline in the world money markets.

Chapter 13
Writing for Localization - Advice for Technical Writers

Darrin Wittwer
Business Development Manager

My career in global language services began in 1990. In that time, I've lived through the incredible advancements in the industry, and in our world, and been touched by some of the tribulations as well. Working with diverse groups of people around the world to remove language as a barrier to business and personal relationships has taught me two important lessons. First, that by working with others I can accomplish far more than, even at my best, I could accomplish alone. And second, life is change … growth is optional. I chose to join Lingo Systems because the company embodies these lessons. Lingo Systems genuinely values long-term relationships over transactions *and* is a catalyst for developing high performance in all organizations. I look forward to collaborating with you.

Technical writers play a crucial role in the product development process. They are responsible for writing the content that describes your products to your end users. Technical writers develop printed documentation, online documentation (such as help files and functional PDF files), and website content. They must take the technical knowledge imparted to them by product developers and present it clearly and concisely to your less technically savvy consumers. As you can imagine, this is not an easy task.

When rolling out your products to the global marketplace, an additional burden is placed on your technical writers. While they are preparing your English documentation for release, they must also keep in mind the requirements for simultaneous or subsequent localization. This process of designing a product so that it can be exported to other countries is known as internationalization.

Documentation that has been properly internationalized is more economical and efficient to localize. Not only does this reduce your localization costs, but it can have huge indirect savings as well. Faster localization often means quicker time to market and accelerated revenue streams.

Some tips that can help your company realize these benefits are described in the following sections.

Layout Issues: Allow for Text Expansion

It is vitally important that your document's layout leaves enough room (i.e., white space) for the likely text expansion that occurs during the localization process. This cannot be overemphasized; formatting the translated document is far easier and more efficient when adequate space is available. Formatting costs can rise dramatically when the translated text must be laboriously manipulated to fit within a cramped space.

As a general rule, assume that your English text will expand 20 to 30 percent when it is translated. This should provide sufficient white space in the English source document for effective localization. In technical documentation, there is a tendency to crowd pages with too much information, impairing the readability of the material presented. Keep in mind that extra white space also makes your English version that much more readable. Because the exact amount of expansion varies by language, please refer to the table presented in Chapter 12 for specific percentages.

Another factor that contributes to text expansion is whether or how hyphenation is used in your document. For example, many German or Dutch words can be much longer than their English counterparts, and many can also be hyphenated. How these hyphenated words are handled will either create an opportunity for convenient line breaks, offsetting much of the extra space that would otherwise be required, or necessitate even more white space.

If your text does expand when translated, you will need to decide whether the localized documents should maintain the same page breaks and the same total number of pages as the English source document. It is generally easier, and therefore less expensive, if page breaks can flow during the localization process. From the perspective of customer support, however, it is often preferable for the localized manuals to match the page breaks in the English version so that support personnel can easily refer to "page 37 of the manual" for solving a problem. If page break matching between languages is necessary, it is even more important to allow for that "extra white space" described above. Matching page breaks from source to target documents can add to the cost of the project, especially when the source document does not allow sufficient white space for text expansion.

When you are ready to hand off your materials to a localization provider, always include a PDF file with the electronic source files. This allows the provider to double-check that the localized files match the electronic file you provided. It is all too easy to accidentally hand off the wrong revision or version of files for localization.

Graphical Considerations: Separate Text from Graphics

Ideally, graphics should not contain text for the simple reason that it eliminates the need to translate it. If text must be associated with a graphic, try to create the text as a separate component in the page-layout application used to create the document (e.g., FrameMaker, QuarkXPress, InDesign). A callout or caption for a graphic should be a text block in the layout program, not an element of an Illustrator Encapsulated Postscript (EPS) file. This requires less work to localize (saving you money), as the graphic text is part of the main document text and not a layer inside the graphic file.

If you must include text in EPS graphic files, remember to leave it in text form. Do not outline the text, as this makes it very difficult and time-consuming to retype and translate.

Screen captures are a special category of graphics. By their very nature they contain text. Translation of screen capture text is accomplished through localization of the software that was used to generate the English captures. Once the software is localized, the screen captures are regenerated. When developing application software, be aware of how the text fits in various

Eva Korinek
Project Manager

My heritage is Czech. Check out this Czech's work ethic!

windows. As with printed documents, avoid packing text too tightly because it, too, will expand when the software is localized. When creating the screen captures, be sure to generate all of them at the same screen resolution and scale, and then save the files in the same format used by the document layout application. You will also want to employ a logical naming convention that will help identify where they are placed.

Reuse Your Content

The cost of localization can also be lowered by including repetitions in your documentation. Most localization providers use tools to help identify text that can be leveraged; that is, once translated it can be reused (at a lower cost) either within the same document or in subsequent versions. (See Chapter 9 for more details about translation tools.) Although you may earn style points for finding new and different ways to say the same thing every time a phrase or concept appears in your English source document, you might also surpass your budget by doing so. Including lots of repetitions in the text will increase the leveraging percentages, which in turn can substantially lower the localization cost, and increase your reader's understanding of the content.

The advent of single-source and content management technologies has everyone thinking about content reuse. For FrameMaker users, it is now possible to structure a document that contains all of your print and online help content and to use single-source publishing tools to publish this content as print documentation, online (HTML-based) documentation, help files, and functional PDFs. Each published output may use all of the original content, or just a subset that can be selected using conditional text. Similarly, XML-based content management systems allow you to store content "chunks" in a database structure and to publish your deliverables from those chunks. Both processes are great for localization because translation has to occur only once for use in many outputs. See Chapters 17 and 18 for more information about single-source content management tools and integrating content management systems. Creating modular content repositories (whether for single-source applications or specialty authoring tools) takes planning to design a logical structure that is easy to reference. Though these tools may offer substantial savings in time, effort, and money, they also require careful planning before starting the process.

Many companies make their documentation available to customers in both a paper-based medium and an electronic form, such as HTML and PDF formats. Other companies have opted to save the printing and distribution charges associated with hardcopy manuals and rely solely on the electronic versions. Both HTML and PDF formats are widely used on the Internet and on alternative media such as distribution CDs because they appear virtually the same regardless of the operating system the customer uses to

view them. For complex, interrelated documents, HTML and PDF formats also offer the advantage of incorporating hypertext, where clicking on the cross-reference, index, or table of contents takes the user immediately to the relevant entry.

When generating electronic documents, the output style should convey the same structured sense of importance that was incorporated in the print document. Moreover, a document that does not use style tags effectively (i.e., one that uses a different style tag each time to produce exactly the same formatting attributes) requires much more time to set up than a document that uses only one style tag to represent a uniform style. Using consistent style definitions throughout your document allows both PDF bookmark data and HTML style tags to be generated in the localized files more easily.

Limit Your Font Types and Font Faces

When selecting fonts for a new document destined for translation, simpler is better. Some languages contain a multitude of accents and special characters that can become illegible if overly ornate or decorative fonts are used. The conventional combination of a standard serif font (e.g., Times) for body copy and a standard sans serif font (e.g., Helvetica) for headings is a good example of font selections that work well for translation. In general, stick to fonts that are clean and crisply drawn, avoiding fonts with exceptionally thin serifs or wispy detail.

Try to keep the total number of fonts used in the document to a manageable number—no more than three or four. Ideally, select fonts that are available on both PC and Macintosh platforms. This facilitates the easy movement of the document across platforms, if required, during localization.

As a general rule, custom or proprietary fonts can be problematic. They are often expensive or difficult to acquire. If required in the final deliverable for branding purposes, expect to provide them to your localization provider. Not only will it reduce time and expense, but you will be sure to get the exact font you need.

Some languages require extended character sets that provide accented letters such as "í." Many specialized fonts do not support languages other than English because they lack this extended character set, so select your fonts carefully. Still other languages need special fonts that are not available as extended character sets. For example, Japanese, Korean, Traditional and Simplified Chinese are considered "double-byte" languages, which means that each written character contains two bytes (16 bits) of data instead of one byte (8 bits). To accommodate these and other character encoding problems, a non-profit organization called the Unicode Consortium created Unicode and related schemas such as Unicode Transformation Format (UTF). Unicode standards have become ubiquitous on operating platforms worldwide, making the display of foreign characters dependable and much easier.

Character styles used in Western European or English layouts are not always transferable to Asian languages. In many cases they are not used at all. For instance, character styles such as bold and italic are not always applicable to Asian type styles. Furthermore, Asian characters do not distinguish between upper- and lowercase. For design purposes, the best way to distinguish Asian characters from surrounding text is to vary the font face or weight (e.g., using a heavier version of a typeface for added emphasis). Your localization provider should offer a variety of techniques to help keep the look and feel you originally intended for your Asian products.

More Tips

Monitor Your Word Count

The cost of localization is directly related to the number of words you write; more words mean higher costs. Monitor your documentation word count by using the "Word Count" command found in your development software. Keep sentence structure and grammar simple and vocabulary choices clear.

Remember Your International Audience

When developing your content, avoid using slang terms and culturally biased graphics. Slang is difficult to translate and understand in a foreign context. Similarly, graphics can also have a cultural bias that can be confusing.

Write Marketing Materials with Localization in Mind

Marketing materials may require special handling as they do not always localize easily. For example, the text and images that succinctly communicate your company or product to an American audience may not be relevant in Europe or Asia. We are all familiar with the stories about product names that take on a second meaning when introduced in another market. If possible, create your marketing materials with localization in mind, and keep the content as precise and globally understandable as possible. If this is not possible, be prepared to provide supplementary materials that will help explain the background, concept, and context behind your marketing campaign. At this point, localization may become less about straight-forward translation, and more about creating the same idea or message while using a different concept altogether. This is where your localization partner will prove their expertise and become invaluable to you.

Internationalize Your Templates

If you use templates and associated scripts to provide a standard look and feel for your layout, it is important to consider localization issues when designing that template. Scripts that automatically capitalize titles, for

example, rarely work correctly on translated content, since capitalization rules vary by language. So keep your target languages in mind, isolating text and automated formatting in clearly identified sections of the template so that your localization provider can easily find it.

Define Acronyms

Write out the full form of each acronym when it first appears in the source documentation. Later, when translated, the first use of the acronym should be both defined and translated in the target language, even if the acronym remains in its English form throughout the document.

Develop a Glossary

Glossaries help linguists understand any industry- or product-specific terms you use in your writing. As you write, keep a separate list of terms that have special meanings. If you provide these terms and their definitions to your localization provider at the beginning of the project you will receive a much higher quality product at the end. See Chapter 7 for more detailed information about glossaries.

Error Message Management

All error messages that appear during installation, operation, or failure of a software application should be stored and maintained in a dedicated location. As new error conditions are exposed, any changes in verbiage between software versions or new messages should be added to the library with appropriate reconciliations from one version to the next.

Software Development—Specific Strings

Any branding or proprietary strings that appear within a user interface should be stored separately within the application. The goal is to collect a single worldwide source for these strings, with localized versions stored separately in language-specific folders or directories. Technical and data-centric strings that do not require translation should be kept with the primary source strings.

Concatenated Strings of Text

Special attention should be paid to concatenated strings of text (individual words linked together in a series or chain, such as "JavaScript" or "nameserver"). In our experience, many translators will assume that multiple words strung together with no blanks constitute single technical terms and will not translate them. Oftentimes, this would be correct, such as with a branded third party product integrated into the application, but it is not always the case.

Ting Fan
Systems Administrator

Dinner: *$50*

Luggage: *$300*

Air ticket: *$1,200*

My wife finally here and us having dinner together: *Priceless*

I got married last year. It took more than a year for my wife's immigration paperwork to be completed. By the time you read this, we are living happily ever after.

Chapter 14
Writing and Displaying Asian Characters

Localizing into Asian languages can present unique challenges. Asian character sets often contain many more characters than Western alphabets; some sets number in the thousands of characters. Have you ever wondered how all of these characters are input and displayed?

In a typical Western font, such as Arial, each character is represented by a single byte of information (or 8 bits), which results in a total of 256 possible characters. This is not nearly enough for many Asian character sets (particularly Chinese, Japanese, and Korean), so Asian characters are programmed as "double-byte"—that is, each Asian character is made up of 2 bytes (or 16 bits) worth of information. Double-byte character sets have more than 65,000 characters available, and the Unicode architecture employs clever mapping techniques, supporting approximately 100,000 characters. This resolves the issue of displaying characters, but it can cause other problems for you if you are developing software that needs to support Asian characters (see "Localizing Asian Software" on page 90).

As for typing Asian characters: You are not likely to find a computer keyboard that contains individual keys for every Chinese or Japanese character; a Chinese keyboard would have to contain over 10,000 keys! Fortunately, some clever methods have been devised to use the standard keyboard. The discussion below is specific to Chinese, but the general concepts are also true of Japanese and Korean.

Entering Chinese Characters

In order to enter Chinese characters into a computer, you need an operating system that supports Chinese character input methods. This could be a native Traditional Chinese or Simplified Chinese operating system, or some other operating system that has either built-in support or third-party software installed for Chinese character input. Once you have the right software, there are three general methods of entering Chinese characters into a computer: typing, handwriting, and speaking.

Typing Chinese characters involves breaking down each ideogram (or character) into a series of alphanumeric characters using a set of defined rules. These rules allow you to create the characters with a standard keyboard. This process is an *input method*. A steep learning curve is required to master the rules of any input method, but these are still the fastest and most effective means of inputting Chinese characters using today's technologies.

Numerous input methods have been developed since Chinese computing was first introduced. Two of the most popular methods used for Traditional Chinese are Zhu Yin and Chang Jie. The most popular input method for Simplified Chinese is Pin Yin. The Zhu Yin and Pin Yin methods break down a Chinese character by how it sounds, representing those sounds with keys on the alphanumeric keyboard. The Chang Jie method breaks down a Chinese character using the character's shape.

Zhu Yin is based on the pronunciation system of 37 sounds and 5 tones that are used in Taiwan. This pronunciation system is familiar to most Taiwanese school children. Chinese characters can be "spelled" with this system. Native keyboards come with the 37 sounds printed on them, so that native speakers can type in Traditional characters. For example, if you want to type the word "Chinese" on an English keyboard, you would type in "5j/" (and a space for the first tone) and "jp6" (6 represents the second tone). If you were using a native keyboard, the Zhu Yin pronunciation symbols would be indicated on the keys. Together, these symbols form the two-character pair that means "Chinese" (中文).

Pin Yin is very similar to Zhu Yin, except that the tone of the character is not considered as part of the input method. Only the component sounds are typed.

Chang Jie, in contrast, is shape-based. It uses 24 familiar characters, each of which stands for a set of related shapes. All but a few Chinese characters can be readily broken up into pieces within this relatively small set of shapes. From this sequence of shapes, no more than five will be selected, by regular rules, to form the code for typing the character. For example, the character 明 is broken into 日 and 月. To follow the Zhu Yin example, you would type the word "Chinese" in Chang Jie by typing "L" for 中 and "YK" for 文.

All of these input methods require the support of your computer operating system. Originally, because Asian characters are double-byte, users had to either use a native operating system or purchase a third-party software bridge for an English system. Today, as a result of Unicode technology, most input methods are supported directly by Windows XP and Mac OS X even on the English version of the operating system. Moreover, since applications such as Microsoft Internet Explorer and Mozilla Firefox support double-byte characters, it is usually easy to write and display Asian languages using Western hardware and software.

Writing Chinese characters on a computer is now also possible, thanks to technological improvements. Various companies have developed Chinese writing pads that connect directly to your computer. Users can write directly on the pad and the software recognizes the handwritten characters and displays them as the appropriate type-written characters on the screen.

Pete Landers
DTP Specialist

I'm a big fan of the local triple A baseball club, and I sing in a barbershop/doo-wop style quartet as well as a male voice choir. Sometimes in the summer I like to take in a little Shakespeare or maybe visit the museums and ballparks of the neighboring states. I think myself fortunate to live in the biggest small town in America. Portland, this rosebud's for you.

For example, at libraries in Hong Kong, a Chinese writing tablet is connected to each computer terminal so that anyone who is not familiar with a standard input method can nevertheless write in Chinese to perform a search in the library database.

Finally, *speaking* can also be used to enter Chinese characters. This method relies on recent advances in speech recognition technology. Users speak directly into a microphone connected to a computer. The software recognizes the phonetics of each word and displays the appropriate characters.

These writing and speaking methods enable users to enter Chinese characters without requiring mastery of the complex rules for standard input methods. Previously, speech and handwriting recognition were only available through third-party software, but now, these methods are supported by the newest versions of Microsoft Office.

They are not without drawbacks, however, as the interpretation of written or spoken characters is far from perfect. The user is generally required to "teach" the software how to recognize his/her style of writing or speaking. Also, these methods are still often slower than the typing methods. As this technology continues to advance, speaking to the computer may one day overtake the traditional typing methods and allow for a more convenient way of entering Chinese characters.

Localizing Asian Software

A significant challenge sometimes arises when localizing your software into Asian languages. Before the advent of Unicode, some software could neither accept Asian character input nor display Asian characters correctly, and some could do one but not the other. To enable the support of Asian (and other foreign) languages, code pages were defined. For example, Windows 95, 98, and Millennium Edition all used code pages that contained 256 code points, where one code point represents one character. For those languages with more than 256 characters, a Double-Byte Character Set (DBCS) was developed.

A major drawback of the code page concept is that a system using a code page can support only one language at a time, since the same code point may need to map to different characters for different languages. For example, under the DBCS system, Chinese can not be mixed with Japanese within the same application unless additional steps are taken.

With the invention of Unicode, these types of issues have been successfully eliminated. However, applications not based on Unicode will require special attention during localization. Read more about Unicode in Chapter 12.

Chapter 15
Localizing Bidirectional Languages

Background

One of the challenges when translating into certain languages is that the writing direction of the target language may be different from that of the source language. Middle Eastern languages like Arabic and Hebrew are written right to left, while Latin languages like English and Romanian are written left to right. When a predominantly right-to-left language contains left-to-right writing (and most do), the result is a bidirectional document. Examples of bidirectional languages include Arabic, Hebrew, Yiddish, Farsi, Urdu, Azerbaijani, Punjabi, Pushto, Dari, and Uigur.

Challenges

Most applications and websites are designed for languages that are written from left to right. As such, applying the left-to-right settings contained in one of these applications to a bidirectional language creates some interesting challenges. Fortunately, there are ways to resolve many of them.

User Interface Layout

The direction of writing affects the way information should be presented and placed. For example, some applications use icons to tell users to go to the "Next" or "Previous" page. Because these icons do not have the same meaning when used with a bidirectional language, however, users often become confused (Figure 1).
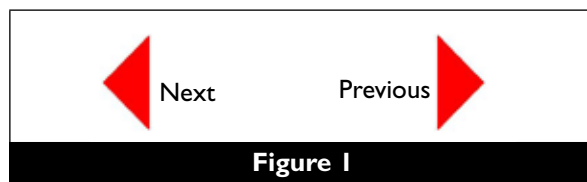
Figure 1

In Figure 1, the "Next" icon for the left-to-right language is correctly understood by the application. For the bidirectional language, however, the "Next" icon is interpreted as "Previous." To fix this problem, the images need to be flipped horizontally (i.e., the one on the right is moved to the left and vice versa) and the underlying functionality must be changed as well (Figure 2).

Oussama Abou-Jamous
Senior Localization Engineer

If you think you can write, try writing in reverse order from right to left. The writing direction of the Arabic script, my native language, is right-to-left so learning English was not easy at all. The most challenging thing for me was to get used to the new writing direction (left-to-right). It was like forcing your brain to think backward to be able to write a useful sentence.
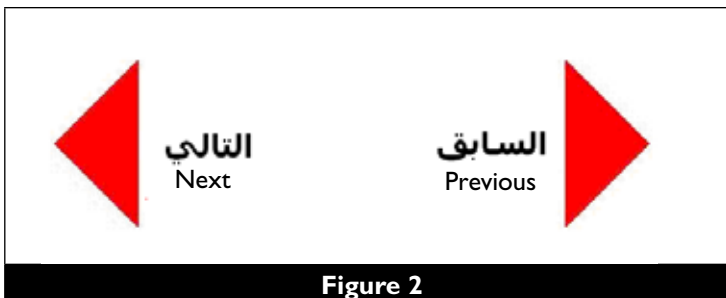
Figure 2

Bidirectional languages can also be confusing to users. When providing step-by-step instructions, you will need to modify the layout for reading from right to left to ensure that the steps are followed in the proper order. Notice in Figure 3 how the following steps will be interpreted differently by users with different reading orientations:

Insert CD in CD-ROM drive	Install application	Click on the application's shortcut to start
Figure 3		

These steps can be easily completed if you read them from left to right, but if you translate them into a bidirectional language and leave them in the same order, the steps will be read from left to right and misinterpreted. The first thing the user will do is to look for the application's "Shortcut to start" without installing the application or even inserting the CD in the CD-ROM drive. Obviously, it is critical to modify the layout of the interface when localizing into a bidirectional language.

Combining Text from Languages with Opposite Writing Directions

When localizing into a bidirectional language, some information, like brand names, does not get translated. A bidirectional document that has text with different writing directions forces the right-to-left text to be stored in reverse order around the left-to-right text. This makes the sentence meaningless and unreadable in most cases. Figure 4 contains a sentence that has a mix of Arabic and English words. Notice how the order of the words is misplaced in the text file and needs to be re-ordered to display correctly.

This is how the characters are stored in a text file.	في البحوثات الخاصة بي Google أنا استخدم موقع
This is how the characters should be displayed.	أنا استخدم موقع Google في البحوثات الخاصة بي
Figure 4	

In most text editors, you can choose the reading order to be right-to-left or left-to-right but this does not fix the issue entirely. It will only fix the reading order in that editor, but once the text is displayed in another application or device, the same problem remains.
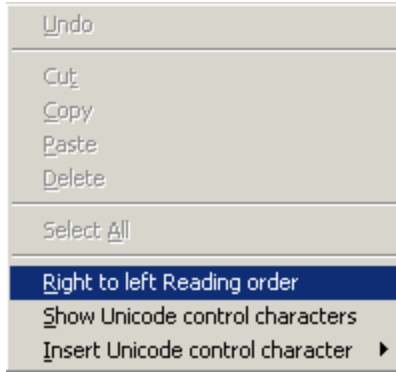
One obvious way to get around this issue is to avoid using left-to-right text (such as English) in combination with right-to-left text (like Arabic). When mixing left-to-right and right-to-left text can not be avoided, the order of the sentence will need to be reversed manually. This will allow you to upload your file to another application and display it correctly.

Alignment and Formatting

Your written content should be aligned to properly display and wrap text at the end of each sentence. If not, the order of the words and the meaning of the sentence will be incorrect. Fortunately, most applications have tools that can change the writing direction of the text. For example, markup languages like HTML have tags you can add to your code to adjust the "DIR attribute" and specify the base direction of text (LTR, RTL).

Most Middle Eastern countries use both Arabic and Hindi character sets for numbers (Figure 5). There is no consistent use amongst bidirectional language users and the choice of which digits to use is often determined by which character set the software supports.

Arabic Digits	0	1	2	3	4	5	6	7	8	9
Hindi Digits	٠	١	٢	٣	٤	٥	٦	٧	٨	٩

Figure 5

Whether you use Arabic or Hindi digits, numbers are written from left to right—with the most significant digits placed farthest to the left (Figure 6). Of course, this is the opposite of bidirectional text, which is typically right-to-left.

European Digit	764
Arabic Digit	764
Hindi Digit	٧٦٤

Figure 6

Punctuation

Punctuation can also be a challenge in bidirectional localization. For example, the exclamation mark, period, and colon are displayed the same way in both Arabic and English; while others, such as the question mark, comma, and semi-colon are displayed differently. In some text editors, when punctuation marks (that display the same) are added to the end of a sentence, they automatically move to the beginning of the sentence unless they are in the middle of a line (Figure 7).

Period at the end of the line is moved to the beginning.	‏. انا استخدم الحاسوب‏
Period between two sentences on the same line displays correctly.	‏انا استخدم الحاسوب. و ابحث‏
Figure 7	

Conclusion

Localization is much more than just translating from one language to another. When localizing into a bidirectional language, you have to adjust the layout of the user interface and reformatting of your application to make it usable.

oops

A marketing slogan gone awry:

HSBC built a marketing campaign around the statement, "Some of man's greatest achievements have been achieved by one simple philosophy, 'Assume nothing'." The tag line, however, was mistranslated as 'Do nothing' in many countries.

Chapter 16

Quality Assurance - How to Be Certain You Got It Right

How many times have you laughed over incomprehensible instructions for connecting a new audio system or assembling a new toy for your child? Intuitively, most consumers realize they are reading a poor translation. But ask yourself: Do you want your customers laughing over something you produced? Most companies seem to understand that the quality of their websites, documentation, and software is extremely important in achieving product acceptance. Time and again we see companies delay projects for weeks while they "wordsmith" their message to get it "just right." When it comes time to translate that message for their foreign markets, however, they choose the cheapest option available. If you are investing in the localization of your product, you should want that product to meet the high quality expectations of your target audience ... wherever they may be. You should also want your product to reflect your carefully crafted image in every marketplace.

Think High Quality from the Start

To maximize quality, consider reviewing your source content and products to determine whether they are "localization-ready." This step begins even before translation, while your materials are still in development! This review identifies possible issues that may affect successful localization. You are probably familiar with the concept of "internationalizing" software (preparing software so that it can be easily localized in the future). We encourage you to apply that same idea to all of your products, including documentation. In our experience, this important step is often overlooked during documentation development, resulting in timelines that compromise a thorough internationalization before beginning the project.

To help make your documentation suitable for localization, your localization vendor should offer to review or "clean up" your English documentation. A grammatically correct and consistent English document greatly facilitates the translation process. Similarly, consistent formatting style (see Chapter 13) significantly improves the document localization process. Keep in mind that if you wait until after the translation process is completed, a poorly formatted paragraph will require fixing in *each* language. If you review and clean up the document before translation begins, the poorly formatted paragraph only needs to be repaired once!

Cristina Tacconi
QA Supervisor

If you count our combined 16 years of employment at Lingo Systems we have worked here longer than anybody. Longer than the company exists, in fact. We just can't stay away from this place! Since we started job-sharing, we love it even more. Now we don't just get to play with languages, we also get to play together!

Thinking about your need for quality assurance and other document localization issues early in the process will dramatically improve your final product. The more thought you put into preparing your products for localization, the more likely you will be to achieve higher quality, lower costs, and shorter timelines.

Quality Assurance

Quality standards are subjective and can vary depending on the use of the final product. With consumer products, you may have exacting quality standards, whereas an in-house training course may have more relaxed quality requirements. The primary quality assurance (QA) standards for localized materials are as follows:

- Translation Quality: Is the source effectively translated into each language?
- Visual Review: Does the document conform to the "look and feel" of the original source materials?
- Functional Testing: Does your software, website, or on-screen display function properly in the localized technical environment?

Translation Quality

Factors that impact linguistic quality include:

- Are professionals or non-professionals performing the translation?
- How many linguistic steps are used—one, two, or three?
- Is there an in-country review?
- What supplementary tools—style guidelines, glossaries, and terminology lists—are used to help complete the translation?

A comprehensive discussion of these topics is presented in Chapters 4 and 5. Each is an important part of the QA process. For example, performing additional linguistic steps to copyedit the work of a translator, or proofreading the draft output for consistency and flow of language, will improve quality and ensure accuracy. Similarly, utilizing your own in-country personnel or representatives to review the final deliverable is another important QA procedure that will add to the overall quality of your translated product.

At Lingo Systems, we endorse each of these steps and recommend that quality assurance specialists verify that the linguists have consistently used country-specific conventions and rules.

Visual Review

Once the formal translation process has been completed, the quality assurance process continues in different forms, depending on the nature of your project. If your project includes printed materials, the QA specialists perform visual validation either on hard-copy or on-screen to ensure that everything in the translated document matches the source document. Among the items typically validated are:

- Completed translation (all items that should be translated are translated, and those that should remain in the source language are not translated),
- Consistent font type, style, and size,
- Correct placement and size of graphics,
- Graphic content (no "clipping" of graphic or text elements),
- Page flow and page numbering,
- Cross-references between text and the table of contents, indices, internal references, screen captures, and/or graphic caption text, and
- Text indentation and alignment.

This list can expand considerably and is normally customized for each project, based on input from the client. To help your vendor develop quality guidelines, provide them with any information that can aid the translation and QA steps early in the localization process. Some examples of helpful information to provide your vendor are:

- Terms and names to remain in English,
- A list of part numbers for your products,
- Measurement units and conversions used in your document (inches/cm, pounds/grams, Celsius/Fahrenheit, etc.), and
- Local contact information for each language (phone numbers, addresses, e-mail addresses, website URLs, etc.).

Functional Testing

Online Documentation

As with your printed documentation, all online documentation should be validated in a QA review. The two main online documentation formats in use today, PDF files and HTML files, also require functional testing. This procedure will confirm that the functionality of the files (e.g., bookmarks, hyperlinks, internal cross-references, etc.) works as intended.

These tests, which are ideally performed by your localization vendor, usually require testing on computers running native operating systems to ensure that the functionality and character displays are correct. Typically, the files are checked for:

- Compatibility with native operating systems,
- Correct display of fonts and graphics using the appropriate browser or reader,

Barbara Bonnema
QA Supervisor

(continued from previous page)

As Dutch and Italian native speakers and linguists, we enjoy working with our mother tongues but we primarily share the leading role in the Quality Assurance Department, where we review many other languages. What keeps us coming back day after day, year after year, is not just our passion for world cultures, but also for wielding our red pen to make sure all documents leave here in perfect shape. (That, and the general excitement that fills the office around the time of the European or World Soccer Cups, of course!)

Jennifer Polis
QA Specialist

Having a background in journalism with an eye for detail can be disconcerting as I go about my day. Awkward sentence structures and misplaced modifiers are enough to drive me batty. This is one of the reasons I enjoy working as a QA Specialist at Lingo Systems so much: I don't have to fuss over whether this or that sentence could be tweaked to flow just a bit better; I can simply appreciate language for its aesthetic value and let the linguists fret over the particulars.

- Correct functioning of hyperlinks, and
- Clear printing of pages.

This list may be customized with other items, depending upon any advanced features that may be added to the PDF or HTML files.

Software and Graphical User Interfaces (GUI)

Software and other user interfaces (UI) also require a thorough QA review. This testing process is similar to PDF and HTML testing. Software should be checked on native operating systems to ensure that character encoding and fonts are correct and that any text expansion does not result in truncated text strings. In addition, hot keys and keyboard shortcuts need to be tested to ensure there are no conflicts and that the keys used actually appear on the local keyboard.

Finally, the application needs to be tested to ensure that the localization process did not introduce any "bugs" (errors) into the software. Typically, software goes through a full regression test to make sure that all areas of the software perform as intended. As with HTML and PDF testing, the actual test may be customized to suit a specific need or client request.

A full discussion on testing and integration can be found in Chapter 11.

Summary

When selecting a localization vendor, look for a partner that cares as much as you do about carefully localizing your product into different languages. A good way to determine this is to inquire about their QA procedures. There are many translators and translation companies in the marketplace, but each one has different quality standards for their work. Quality assurance steps must be performed at each stage of a project in order for a localization vendor to meet your expectations for the final product.

oops

On a Tickle Me Elmo box in China:
Crap your hands

Chapter 17

Single-Source Content Management Systems

by Paul Trotter

More and more businesses are expanding into non-English speaking markets. A critical success factor for this expansion is high-quality, cost-effective, and timely-translated written content. Responsibility for this typically falls on internal translation departments or localization partners. Though translation is often an afterthought, it can come at a high price, in some cases exceeding the cost of writing the original content after only a few languages.

Current approaches to localization rely on technologies and processes that have minimal scope for improvement. The localization industry is under increasing pressure to find new ways to improve cost-efficiency, quality, and time-to-market.

In this chapter we will try to explain what content management is and how it can help your organization write higher quality and more effective documentation more efficiently. We will also discuss how to reuse and share content across documents, have strict control over standards and branding, publish content to print, help, and web formats, and reduce the cost of localizing your content.

What is Content Management?

So what is content management? The first thing to say is that there is no single agreed definition. Content management is a relatively new discipline, and if you ask the many suppliers of content management software they all have different definitions. Of course, most of them make the definition suit what their software does.

It is fair to say that most people regard content management as applying solely or mainly to the management and delivery of web content. This is a very limited view. Content management software covers a much wider area and can be categorized as follows:

- **Web Content Management**—This was the first and is the most common use of the term "content management." These systems are primarily used to help manage websites and web content. In this context the word "content" refers to any resource used to build a website. Most of these applications are only concerned with managing the delivery of the website. The authoring and maintenance are done by other products.

Paul Trotter
CEO,
Author-it

Paul is the founder and CEO of Author-it Software Corporation. Paul has been involved with documentation for more than 15 years.

- **Document/File Management**—Document and file management systems are designed to manage whole documents and other files rather than the words and pictures inside them. They know little about what the files contain and treat them as just a "blob" of data. They rely heavily on users defining and applying metadata to give them more information. In practice, however, metadata is often not applied and the applications are little more than filing systems.

- **Digital Asset Management**—Very similar in nature to document and file management systems, in that they manage files; but because they focus on multimedia there is little or no functionality for text-intensive files. These applications are used mainly to create a central repository for graphics, video, flash, and other multimedia files, and provide archive, search, and retrieval functions.

- **Source Control**—Again, similar to document and file management systems but they are primarily concerned with managing source code, which are pure text files. They usually have poor support for dealing with binary content and often provide integration with software development environments.

- **Enterprise Content Management**—This is perhaps the fastest growing category in content management and does not yet have a clear definition. Most providers in this space are actually combining many of the other categories and calling it "Enterprise" as it provides a wider scope.

- **Single-Source Content Management**—These systems provide the most benefits for localization. Rather than storing documents, they store and manage the content that is used to assemble those documents in small reusable components. These components can be anything from a single word to many paragraphs, or other components like graphics or links.

Single-source content management is an overall process for originating, managing, and publishing content right across the enterprise and to any output.

Content management should be an end-to-end solution providing the ability to track, manage, and control what happens to your content at all stages in the documentation cycle: from authoring and importing, to storage and document assembly, to multi-output publishing.

What is the Difference Between Managing Content and Managing Files?

The answer to this question is the key to why single-source content management provides so many benefits over traditional file management systems.

The most important aspect to managing any kind of data is to control how it is created and changed. This is the cornerstone of enterprise applications of all types and is the only way you can truly manage information. The next step is adding value to it.

The traditional approach to document and file management has been to move the files from the file system into a database where they are stored in exactly the same format as they were created. These systems typically provide access control, versioning, metadata tagging, and search capabilities. There is little control, however, over the modification or creation of the files in the first place. Instead, they rely entirely on other applications to do that.

Let us look at this problem from a different perspective. Let us say your organization is using Excel spreadsheets to manage their financial accounts. At some point this approach becomes unmanageable for a variety of reasons. It is decided to move to a purpose-built accounting system that uses a backend database, allows multi-users, provides audit trails, has financial reporting, and is able to manage the underlying data.

Would you just move the Excel spreadsheets as they are into a file management system and expect it to magically create a profit and loss statement, or chart of accounts? Of course not; that would be impossible.

the accounting evolution
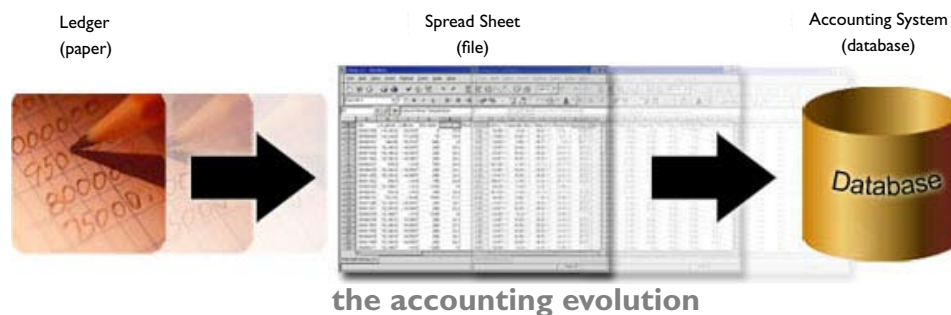
Instead, you would move the data from the spreadsheets into the predefined relational database structure provided by the purpose-built accounting system. Now you would be able to get all your reporting, ensure data was entered correctly, have multiple users editing without fear of overwrites, and exercise a much greater degree of security over your data.

Would you expect to be able to continue editing your accounts in Excel? Of course not—the information is no longer in Excel format, and doing so would bypass your controls and auditing. You would now edit the information in a controlled fashion in the accounting system. No longer would you get an unbalanced transaction or have information changed by unauthorized sources. Best of all, your reporting is a mouse click away.

Single-source content management provides the same evolutionary leap for content. It provides a more effective and more efficient way of authoring, managing, publishing and localizing your organization's documents, images, web content, etc.

Why do You Need Content Management?

Content is an Asset

Generating content takes time and money—often lots of both. So content should be treated as the valuable asset that it is.

To get maximum value from your documentation resources, you should be able to do a number of things:

- Reuse content across documents without copying so that you can write it once and maintain it in a single place, no matter how many times you have used it,
- Use content created for one purpose equally well in other contexts and for other purposes,
- Translate reused content once and have it automatically reflected wherever it is used,

the content management evolution

| Typewriter (paper) | Desktop Publishing (file) | Single Source (database) |

- Publish to print, help, and web outputs without having to modify or make different versions of your content, and
- Involve more people in the documentation process, such as subject matter experts, application developers, localization teams, and trainers.

These features have the potential for increasing the quality and consistency of your documentation, reducing the cost and time involved in producing it, and gaining more value from every piece of content that you create.

Control is Essential

Assets are of no use if you can't manage them. Having tons of content that you cannot find, organize, protect, or use effectively is simply a waste of time.

Involving more people is a good idea, but it requires serious organization. Wider access can be a disaster if the system can't cope.

To properly control your content, you must be able to:

- Set and enforce your standards to ensure consistency and quality,
- Control who in the organization can create, see, and use content,
- Find the content components when you need them,
- Manage the content life cycle through drafts, reviews, localization, release, and archiving, and
- Control what can be published to each output channel and by whom.

What are the Savings and Benefits?

An Example of Cost Savings

Whether you choose to manage the translation in-house or to outsource it to an external vendor, localization can be a complicated and expensive process. On the first mention of localization the immediate reaction from your financial department may be to reach defensively for their wallets. Costs can be unpredictable and can quickly get out of control, particularly if you don't know what to expect. Let us look at an example to put this in perspective:

The average cost a translation agency will charge for translation, editing and layout is around $0.35 cents (US) per word. Take a document with 500 pages and an average of 200 words per page. That's 100,000 words, so you are quickly looking at *$35,000*.

Now remember that is just for the initial translation. There will be more costs when you make modifications to the original document and need it retranslated. Most translation agencies use translation memory tools which help reduce the effort involved in retranslating a document, but they still charge for the whole document (albeit at a reduced word rate for the text already translated).

When using translation memory tools, a *fuzzy match* is returned when a text segment is similar but not identical. An *exact match* (100%) is returned when there is no difference or variation between the two segments. Translators often charge different rates for exact matches, fuzzy matches (with the match falling between a certain percentage), or new previously untranslated text.

Let us get back to our example. You now modify 5% of these pages, and add 20 new pages. Without allowing for *fuzzy matches*, the cost of retranslation can quickly approach $11,000:

20 new pages - 4,000 words @ 35 cent per word	$1,400
5% change - 5,000 words @ 35 cents per word	$1,750
95% unchanged - topics with 95,000 words @ 8 cents	$7,600
Total cost of update	**$10,750**

Over time, these costs quickly mount up. Our example was just one document into one language. Translate that same document into 10 other languages, and multiply the cost 10 times. Translate a further 10 or 100 documents into multiple languages, and watch your costs skyrocket!

How Single-Source Content Management Reduces Translation Costs

Using a content management system that stores and manages content in XML format can facilitate localization. It can also yield significant savings.

1. You only translate objects that have been modified.

   For example, let us go back to our 500 page document which we have now updated. Rather than sending the translator all 500 pages again, only the 20 new pages and the 5% of modified pages are exported as XML. Using our previous example this would reduce the cost of retranslation from $10,750 to $3,150!

2. Text is only translated once.

   The same components are reused in multiple documents. For example, the same copyright notice (or even an entire introductory chapter) may be used over and over. Each component only requires translation once.

You can even reuse content as small as a phrase, sentence, or paragraph which takes reuse even further, and again, only pay for translation once.

Cross references and hyperlinks do not even require translation. Because they are inserted at publishing time, taking their text from the heading of the component they reference, they are not stored in the text, resulting in less to translate. Likewise, reference text such as "See" and "on Page" is defined by templates, so only the template requires translation.

Our studies have shown an average 30% reduction in word count through reuse of content.

3. The XML files do not contain formatting.

   When the same text string is found using different character formatting, memory translation tools do not always identify it as an exact match. Because the XML files in a content management tool do not contain formatting, this helps increase the exact matches found.

Benefits for Localization

When you manage your content at a more granular level there are a number of things you can do that cannot be done with whole documents. Some of the specific benefits to localization are:

- **Translate Content Once**—The system knows what content is translatable, has been previously translated, is reused, or has been added or changed since the last translation. Only content that actually requires translation is sent to translators, which significantly reduces word count and cost of translation.

- **Faster Time to Market**—Localization and content creation can run in tandem, allowing translation to finish much sooner. Content is created in small discrete components that can immediately be sent for translation. This avoids the costly exercise of translating drafts or waiting for completion of the entire source content.

- **Automated Single-Source Publishing**—Once source content is translated and reviewed, it can be published directly to print, help, and web formats without tweaking or rework. This provides substantial savings, and eliminates inconsistencies in translation across delivery formats.

- **Cleaner Translation Memory**—Translatable XML contains only text and semantic markup, increasing translation memory accuracy, and eliminating the effect of formatting on memory matches.

- **Improved Accountability**—Only content that requires translation is sent for translation. Each piece of content has an accurate word count and is known by all parties in the process, avoiding any surprises or disputes.

Overall Benefits of Content Management

Single-source content management provides significant benefits and cost savings over traditional document authoring and maintenance methods. Some of these are:

- **Faster Time to Market**—Because authors spend far less time creating and recreating the same content, reviewers spend less time reviewing and translators spend less time translating. Publishing to print, help, and web formats is fully automated. This is achieved by controlling standards, eliminating duplication, and effectively managing creation, localization, and publishing of content.

- **Efficient use of Resources**—By eliminating repetitive creation and maintenance, more of your resources can be devoted to improving the quality of the content and adding value to your documentation. Many clients report savings in excess of 20% through reuse of content.

- **Major Production Cost Savings**—Efficient creation, maintenance, and management of content will naturally result in major cost savings. You achieve more documentation for less outlay and the time taken to produce a page through traditional authoring tools can be halved.

- **Lower Translation Costs**—Content is translated only once no matter how often it is reused. Translators only ever work on new or changed source content, so you do not pay for them to handle unchanged text. Real projects have shown reductions in translation word count in excess of 30%.

- **Improved Quality and Usability of Content**—Through enforcement of standards you can guarantee consistent documentation structure and formatting, increasing readability and usability. Using single-source content ensures 100% consistency wherever it appears.

- **Improved Workplace Satisfaction**—Freeing authors from tedious, time-consuming tasks such as formatting and repetitive updates allows them to concentrate on creating and improving content. Reviewers gain by reviewing content only once, regardless of the number of end deliverables. Writers save 95% of the time they usually spend formatting content.

- **Increased Customer Satisfaction**—Consistent, accurate documentation of all types means fewer calls to customer support because you are providing the right information, at the right time, in the right format.

Chapter 18
Integrating Content Management Systems

For many years now, both machine translation and content management systems (CMSs) have been "the next big thing" in the localization and technical communication world. Unfortunately, the complexity and limitations of the underlying technology, coupled with prohibitive costs, have limited widespread acceptance for all but the largest companies. As is often the case with new technologies, however, the newest generation of tools are more affordable, the functionality more approachable and the barriers to entry lower.

The other reality of next generation technologies is that users find new ways to leverage them as the tools evolve. In the localization business, many companies have discovered that CMS tools not only reduce the cost of authoring and publishing their English content, but that even larger savings are available when they translate their content into other languages. The biggest savings, however, result from integrating a CMS with both your own business systems and your localization vendor's translation management system (see Chapter 19) to produce a fully automated localization workflow.

The current state of machine translation is covered in Chapter 20.

How does a CMS reduce localization costs?

So just how does a CMS reduce localization costs? If you consider a traditional localization process from start to finish, the answer becomes apparent. Your localization vendor does a lot more than simply convert your source language content into a target language. Because they are also responsible for ensuring that the target language content has the same look and feel as the original English source document or web page, they are an integral part of your publishing process as well. In fact, although many people do not realize it, localization often requires the use of language-specific publishing templates, fonts and character sets, and may even require selective publication of content for individual regional markets.

With a CMS, your localization vendor will no longer need to perform desktop publishing on your target language deliverables. Instead, this function is performed within the CMS itself. Your vendor will also have less file preparation and localization engineering to perform since the format tags that dictate how your text will appear on paper or on screen are no longer comingled with the text itself. There may even be savings from shorter timelines since content can be localized before it is published into a final document.

Joseph Starnes
Solution Architect

Joseph Starnes. Solution Architect. What does that mean? That's a very good question. Another good question is, "What will be the true cost of this faux leopard jacket if I purchase it with my credit card and then only make the minimum payment each month towards the $1,723.89 principle?"

"Wow! That much! Really? Haha. Maybe I don't need another jacket just now."

And it's just that sort of sharp thinking that Mr. Starnes brings daily to Lingo Systems. Weighing every option, sometimes twice, occasionally with a scale. Shifting gears, nimbly, sometimes even using the clutch.

Reusing translated content is the best way for your vendor to save you money, however. This topic is discussed in detail in Chapter 9, but suffice it to say the best illustration may be to liken it to a translation memory on steroids. At Lingo Systems, we have a customer who recently implemented a CMS and now publishes their online Help from their print docs. Because they only pay to translate the content once, they now get their help translated for *free*!

Is a CMS worth the trouble?

In the context of localization, there are several business conditions that indicate it might be time to look at adopting a CMS. Of course, there are also many non-language reasons as well, but that is a different discussion.

- Frequency of translations: Is it likely that you will translate your source content?
- Quantity of content produced: Does your writing group create and maintain tens, hundreds, or thousands of pages of documentation?
- Frequency of updates: Once content is translated, is it subject to change on a frequent basis, or does it remain stable?
- Number of target languages: Do you consistently translate into more than one target language? Do you choose to translate some content into some languages and not others?
- Published outputs desired: Do you publish the same content in multiple formats (e.g., a spec sheet for print, web, and specification summaries)?

If you answered "yes" to any one of these questions, it may be worth your time to investigate whether a CMS is right for your company. When making this evaluation, you will also need to determine whether a general CMS strategy or specific CMS tool is appropriate for your business. Some of the factors to consider include:

- The structure of the writing teams,
- The content creation, editing and approval process,
- The publication process, and
- How to migrate your existing English and localized content from its current format into a CMS (importing legacy content and setting up publishing templates can be a major undertaking).

As with most business decisions, however, perhaps the most important factor is one of economics. Implementing a CMS can significantly reduce your localization costs. One Lingo Systems client is now able to translate three times the amount of content for each dollar spent on localization. If you plan to manage a significant amount of content in multiple languages, the resulting return on investment (ROI) of a CMS may make the decision a "no-brainer."

Manual Writing vs. a Structured CMS

Moving to a CMS requires significant changes to your writing and publication process, as well as elemental differences in how you will manage content in the future. Unlike traditional writing, with a CMS the writing and managing of content is independent of the publishing process. This means that in many cases the content being written will be used in more than one place, and possibly published in more than one type of media.

The key to meeting this challenge is structured writing. Instead of producing a sequential document that starts at the beginning and moves from one chapter to the next, writers produce independent "chunks" of content that are then stored in a relational data base for subsequent use.

An analogy is the evolution of automotive manufacturing. Before Henry Ford, each car was essentially hand built from individual parts. Workers started with a bare chassis, and put together the car part by part until it was finished and driven away. As manufacturing techniques evolved, fabrication shifted from individual parts to major assemblies produced all over the world. Today, most cars, even the traditionally hand-built models, are constructed from dozens of pre-manufactured components resourced worldwide. Most significantly, the major frameworks are often used in more than one model, and even brand, of car.

Structured writing is a similar concept; contributors all produce content that is akin to an automotive assembly line. At the time of publication, the appropriate content is collected into a final product that is then plugged into a template to be presented in whatever fashion is desired.

Factors to consider when selecting a CMS

- When implementing a CMS, there are many contributors, constituents and downstream consumers who will likely see the nature of their responsibilities change. In our experience, some of the parties whom you may want to involve when evaluating a CMS are:
    - Technical writers and other content contributors,
    - Localization manager or coordinator,
    - Webmaster,
    - Marketing managers,
    - Internal IT Service and Network Engineers, and
    - Localization Vendor.
- ROI is a popular means to measure the potential success of just about any technology application. In terms of a CMS and the resulting downstream impact on localization, we suggest you define ROI over a 6-, 12-, or 18-month horizon, and evaluate payoff in dollars and production time. If you have a consistent basis for comparison, calculating the ROI of the acquisition cost against your localization

Michael Phoenix
Software Engineer

Most of my career has been in software development and I'm really excited to be working in Web applications. It's been a goal of mine since receiving my masters in Computer Science a couple of years ago. It's also neat to be working at a prominent translation and localization company. Over the years I have had many opportunities for international travel and have come to enjoy associating with people from different backgrounds and cultures. Being an employee here allows me to do that and to help companies succeed in the global marketplace.

cost is a simple matter. It is even easier if you have your prospective localization vendor do the calculations for you. The CMS tool vendors may also be able to generate an ROI calculation, but we believe it is best to have a custom calculation developed based on your particular business situation.

- Depending on the tool you choose, it may be advisable to run a pilot project as well. If you take this route, begin with a round of translation in order to test the entire workflow. You will also want to get an idea of just how much you can leverage your existing translation memories against the new content format. You may lose some leveraging between your traditional file formats and the CMS content as a result of conversion issues.

- Driven largely by the need to manage enormous volumes of constantly changing web content, commercially available CMS tools have sprouted up everywhere. Be careful—these tools were not all created equal. If you are certain that you only want to manage your web content, there are a vast number of CMS tools to pick from. If you need to manage both web and print deliverables, however, the choice of tools and methodologies narrows significantly—even more so if you require an out-of-the-box solution.

- CMS tools significantly affect localization workflows. As such, you will want to be very deliberate about how you manage what content is routed for translation and when. For example, once you have translated one block of English content into six target languages, the means by which you maintain a relationship between these equivalent chunks becomes essential. Also, if content is planned for publication in multiple places, how you manage updates and maintenance in each location is equally essential. If you think that managing your English content in a CMS structure is a feat, the reality of managing that content plus six localized versions will really get your attention. Fortunately, some CMS tools make this process much simpler than others, and a few make it downright easy.

Summary

CMS-based publication and localization has progressed from the early-adopter, trail-blazing phase to a mainstream publishing activity. If your organization is willing to fundamentally change how it creates and publishes content, a CMS may be the right move.

CMS-driven time-to-market improvements, writing productivity, and *dramatic* savings in localization costs are all well documented and very real. Also significant are the gains in consistency in the look and feel of your content and branding.

It is difficult to overstate the promise of content management systems; these tools are changing the very nature of content creation, publishing, and localization in the same manner that translation memory tools did a decade ago.

Chapter 19
Translation Management Systems

The localization industry services companies of all sizes in all sectors. The result is that localization buyers have very different needs depending on the complexity and volume of their translatable materials, as well as their short- and long-term strategies for growth in both local and international markets. In other words, every company has its own methods and processes to communicate with its multilingual workforce and customers. For this reason, LSPs such as Lingo Systems must have flexible processes and tools to provide each customer with the best solution to meet their needs, keeping in mind that the "best solution" is not always the most expensive or the most complex, but rather the one that provides the desired output quality at the best possible price.

In our experience, smaller consumers of localization services are mostly interested in obtaining a good quality product using an easy, straightforward and manual process where files are exchanged via email or FTP. This simple approach is most appropriate for projects that are few and far between. On the other end of the spectrum, however, are large multinationals that need to submit projects from, say, five different divisions in 40 different countries and also monitor the status of every project, wherever it originated, 24/7. These larger consumers may also require integrated server-based solutions for TM and terminology management, centralized management of all data, automated integration with CMS tools, or customized localization workflows.

Several LSPs, including Lingo Systems, have developed proprietary software tools (SAAS, off-the-shelf, home grown, or even open source) called translation management systems (TMSs) to provide these complex services. Although TMS platforms are a fairly recent addition to the localization world, it is a segment of our industry that is moving at a very fast pace. TMS growth is driven in part by the increasing use of content management systems on the client side as well as the ever-evolving expectations of our customers. In a world where anyone can perform any task online, from the most benign to the most complex, at any hour of any day, it is normal for a customer to expect the same level of service from an LSP.

In an effort to recover development costs, some vendors try to indiscriminately put all of their customers in the TMS box, but this solution should really be reserved for clients with specific needs. These include users who need a fast turnaround on their localization projects, global companies who have a great many users requiring access to a single localization resource, and those with internal enterprise systems that can integrate translation.

Richard Johnson
Systems Architect

Systems Architect is our term for "technology dude." I look at nearly everything we and our clients do and need, and then try to figure out how we can use technology to make that happen faster, more accurately, and with less disruption. It's a wonderful work environment for a scatterbrained mad scientist. Boredom is no issue; It's often a full-time job meeting the wildly divergent day-to-day needs of clients. At the same time, Lingo gives me enough freedom to investigate new opportunities and approaches, and exhaust the 99 hair-brained ideas that don't work in order to find the one that does. In my spare time I work on Lingo*NET*. If you don't already use Lingo*NET*, consider it. It's a wonderful tool that might well save you time, effort, and grief.

If your organization could benefit from a TMS, keep in mind that each one is different and there is no one generic TMS. Every vendor offers different features and functionality, and some charge for these tools while others, like Lingo Systems, do not. Thus, when you shop for an LSP, ask whether they have a TMS, what capabilities it has, and how much you will have to pay to use it. Keep in mind that since a TMS will reduce your localization vendor's project management costs, it is reasonable to expect a free solution if you bring enough revenue to the table. The more common features that you are likely to find in a TMS are:

- Project Estimation: Be sure to confirm that your core file formats are fully supported in all stages of the workflow.
- Project Management: Delivery date, schedule, project specifications, etc.
- Financial Information: Leveraging analysis
- File Management: File delivery, archival, etc.
- Reporting: All information that you need to know about your localization projects should be available.

The best systems offer more advanced features to make localization workflows even more ubiquitous, faster, and transparent, while reducing costs even further. These include:

- Systems Integration: Full API to create connectors to content management systems, business management systems, translation assets management software.
- Resources Management: Gives you the ability to assign specific resources based on the type of project being uploaded.
- Workflow Customization: Where the user can decide which steps the project will go through based on project requirements.
- Machine Translation Integration: This is not a standard offering, as MT technology is not yet ready for primetime and always requires human intervention to ensure the quality of the output. But it is clear that the localization industry is moving in this direction, and machine translation can already be used as a free service on many portals.
- Live Chat: Audio and video content is one of the fastest-growing segments of the localization industry.

The sky is the limit when it comes to the features that could be offered to gain even more efficiency in the localization process and reduce costs. Unfortunately, it can be very difficult to sort through all the TMS applications currently on the market, and some companies are not always forthcoming when it comes to releasing information. If you are interested in obtaining more detailed information on different TMS tools available, you can visit the Common Sense Advisory website where subscribers have access to a full TMS report evaluating different offerings (www.commonsenseadvisory.com).

In summary, if you believe that this technology will benefit your business, it is important to focus on what you want to achieve with a TMS rather than selecting the system with the most bells and whistles. You should also remember that a TMS is just a gateway to your translation resources; there are still human beings performing the work behind the curtain. Don't be fooled and always test the quality of the output rather than the design of the TMS interface.

Chapter 20
Dispelling the Myths of Machine Translation

by Uwe Muegge

reprinted with permission of tcworld magazine, copyright July 2008 www.tcworld.info

It is not surprising that myths, half-truths, and misunderstandings abound regarding machine translation: It seems as if the experience most players in the translation field have with this technology does not go beyond toying a little with one of the free online translation tools. Almost every week, I come across an article informing its readers either that machine translation is and always will be a complete waste of time or that machine translation, while being a waste of time today, might actually be useful some time in the distant future. In the hope of setting the record straight, here is a closer look at some of the most common myths about machine translation.

Myth: Machine translation simply does not work

With free online translation services available all over the Web, anyone can run a text through a machine translation (MT) engine and then share the results with the public as proof of the fact that machine translation is capable of little more than the most rudimentary rough translations (gisting), and, of course, providing nearly endless entertainment.

The main problem with these 'tests' is that using any of the free online translation environments gives only a glimpse of the true power of a full-fledged professional machine translation system. For example, the typical online translation service does not allow users to select a subject field or provide user terminology, let alone set stylistic preferences. In fact, many—if not most—of the free text translation tools support no translation parameters other than the specification of the language pair and the source text. No wonder that the translations these machine translation websites produce can be so ridiculously off target.

Fact: Machine translation improves the productivity and consistency of human translators
Whenever new source text for a project is created, that text will have to be translated at some point. Even when you work in what is considered a state-of-the-art globalization environment, i.e. an integrated content management/translation workflow system, you will end up with a certain percentage of low match/no match sentences.

Uwe Muegge
Corporate Terminologist, Medtronic

Uwe Muegge is the corporate terminologist at Medtronic, a manufacturer of medical technology. He serves in ISO Technical Committee 37 SC3 Computer Applications in Terminology, and teaches Terminology Management and Computer-Assisted Translation at the Monterey Institute of International Studies in Monterey, California.

info@muegge.cc
www.muegge.cc

In a well-planned and well-managed globalization project where writers, as well as software developers, use a comprehensive project glossary and a style guide aimed at easy readability/comprehensibility, the low/no match sentences can be pre-translated in a machine translation system before being edited by human translators.

The benefits of machine-generated pre-translation include:

- Translators always have a proposal to work with instead of starting each new translation from scratch. A representative case study recently conducted at Symantec indicates that the productivity of human translators can double when unknown sentences are pre-translated in a machine translation system.[1]

- While most translations will require some editing and many even rewriting, it is fair to expect that a considerable percentage of machine-generated translations turn out to be perfect (this is especially true for short instructions, headings, legends, and the like).

- At a minimum, key terms will be translated correctly and consistently. And not only that, in most cases these terms will also be inflected correctly and appear in the correct singular or plural form. (Try to do that with your translation memory!)

Fact: Machine translation enables the translation of material that would otherwise not be translated

Very few organizations, if any, currently translate all materials into all the languages spoken by all of their current or future customers. The primary reason for this is that for many types of documents, especially in the after-sales domain, the budget is simply not available for large-scale human translation.
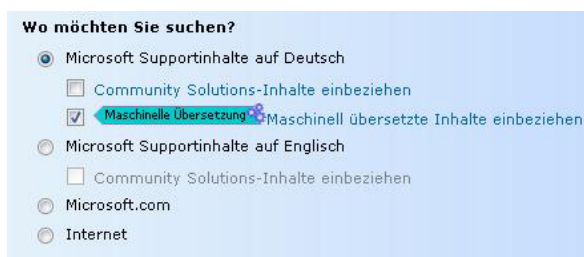
*German search page for the Microsoft Knowledge Base
with machine translation option enabled*
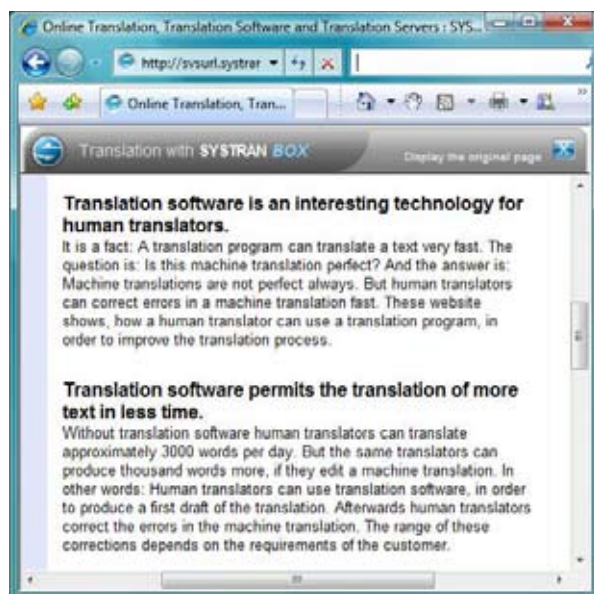
Example of a German > English machine translation from the author's website, www.muegge.cc

A number of organizations are using machine translation solutions for making large volumes of text available to their global customers in their local language without involving any human translators in the process. The Microsoft Knowledge Base, which contains more than 200,000 documents in English, is a well-known example of a text repository where the number of machine-translated documents by far exceeds the number of those translated by humans.

Myth: Machine translation systems can only handle word-for-word translation

The belief that machine translation is basically limited to the sequential substitution of words in the source language with words in the target language is as widely-held as it is wrong. All popular machine translation systems, including the free online translation services such as systransoft. com, translate.google.com, and windowslivetranslator.com employ highly sophisticated algorithms that are the result of years of research and development.

Fact: There is not one but many very different machine translation technologies that are all capable of producing excellent translation results in the right environment
Machine translation has been around for more than 50 years, and during this half century a wide range of MT technologies have evolved (e.g. dictionary-based, rules-based, example-based, statistical) plus countless hybrid forms. Here is a brief discussion of the three machine translation technologies that are most relevant for commercial applications today.

Rules-based Machine Translation

Rules-based machine translation, also known as transfer machine translation, is the dominant MT paradigm today. SYSTRAN, Babel Fish, and @promt, to name just a few, are all rules-based systems. Rules-based MT systems use a three-stage translation process:

1. Analysis: Parses the source sentence to create a tree of the syntactic structure of that sentence.
2. Transfer: Converts the syntactic tree for the source language into the corresponding tree for the target language.
3. Generation: Populates the target tree with corresponding words to create a sentence in the target language.

The benefits of rules-based machine translation include:

- Mature, proven technology that can be implemented quickly and at relatively low cost.
- Many commercial systems available, covering many language combinations.
- Highly customizable through dictionary and style settings (some systems also support the customization of the rules base).

Rules-based machine translation systems have been in use in commercial settings for many years, e.g. at Autodesk, Daimler, and the European Commission's Translation Service. The two primary challenges for rules-based MT are first, that the rules base of any system is by necessity limited, meaning that for best results, authors need to adjust their writing style; and second, while commercial rules-based machine translation packages are available for dozens of language combinations, many languages are still not covered.

Statistical Machine Translation

Statistical machine translation (SMT) is getting a lot of media attention these days, especially after Microsoft announced that it is using a proprietary SMT system to translate its huge Knowledge Base document repository,[2] Google won a large-scale machine translation evaluation contest with its statistical machine translation engine.[3] Statistical machine translation systems typically consist of two major components:

- Translation Model: Generates translation proposals based on corresponding word sequences in aligned source and target training data.
- Language Model: Selects the best translation proposal based on training data in the target language only.

The good news about statistical machine translation is that once an SMT system has been trained on customer-specific data, this is the MT technology that typically produces the highest translation quality. On the flip side, that training effort requires a substantial body of existing translations: Language Weaver, the leading vendor of statistical machine translation systems, recommends a bilingual corpus of two million words or more per language pair. Because of the demanding training requirements,

combined with the fact that statistical machine translation systems tend to have a higher sticker price than some of the rules-based systems, this MT technology is primarily used by government agencies—the intelligence community in particular—and large corporations.

Direct Machine Translation

In its most primitive form, the only thing a direct machine translation system does is to replace the words in the source language with words in the target language in the same sequence and without any linguistic analysis or processing. The only resource direct machine translation uses is a bilingual dictionary, which is why this MT technology is also known as dictionary-driven machine translation. Due to this rather unsophisticated technology, direct machine translation has been considered obsolete for many years, and there are hardly any commercial products available that use direct MT.

Despite its limited capabilities, I strongly believe that direct machine translation still has a place in today's arsenal of automated translation tools. For a number of common real-world applications, word-for-word or phrase-for-phrase substitution is all that is required for successful translation. Think of domains where both vocabulary and syntax are standardized, as is the case with weather reports, financial profiles, and many e-commerce applications.

In one recent implementation, Medtronic, a large medical device manufacturer, used direct machine translation to translate a large product database into multiple languages.[4] Human translation was not an option for this project because of cost and, yes, quality concerns (an analysis of previous human translation projects indicated an unacceptably high error rate among numeric values such as product numbers and dimensions). Also, initial tests had shown that both translation memories and rules-based machine translation systems produced poor results with text that has the following characteristics:

- little or no repetition on the sentence level;
- high repetition on the word/phrase level;
- telegraphic/elliptic style, e.g. 'winds from southerly direction, speed reaching 55 km/h', 'American Technology Associates (AMTA) strong buy, Avion (AVIO) market outperform', or 'plate 2456dr15 right-angled, slotted, 15 ea'.

This type of translation project is most definitely among those that any self-respecting human translator could easily do without. And since direct machine translation does not require human post-editing in a best case scenario, using MT in this kind of environment might for once be welcomed by translators (who would hate to do these translations themselves) and translation buyers (who would love the idea of almost instant, almost free translations).

Myth: Machine translation is only for large organizations

Yes, it is true: If you read any success stories about machine translation, they typically come from the Caterpillars, Microsofts, and Symantecs of this world. But that is true for many—if not most—emerging technologies. It is also true that some of the most powerful machine translation systems in use today are the result of the multi-million dollar research and development programs only corporate giants can afford. But that does not mean you have to spend big bucks to deploy a machine translation solution.

Fact: Being both affordable and user-friendly, many machine translation packages are available for even the smallest of businesses, including freelancers

If you do a little research, you will find that many commercial machine translation packages are in the same price range as their translation memory counterparts, and that is mostly true for both workstation solutions for single users and client-server solutions for many users. And the secret is out that while corporate and small-business versions may differ in many ways, the core translation engine is typically the same in both products. In other words: In terms of out-of-the-box translation quality, there is generally little, if any, difference between the $1,000 professional version and the $50,000 corporate version of a given machine translation product.

In addition, the developers of commercial machine translation systems have invested heavily into making their products as intuitive to use as possible. In fact, I would even say that it is easier—and certainly faster—to produce your first translation with a typical MT product than it is with the typical translation memory tool.

A few more facts to consider:

- Many low-priced machine translation products feature a built-in translation memory (TM) module to improve the efficiency of the post-editing process ('never correct the same mistake twice'), and a few MT systems (like @promt Expert) offer seamless integration with the SDL Trados translation memory system.
- A number of translation tools vendors (such as Across) that cater to small and mid-sized companies offer TM-MT system bundles and/or MT integration via API.
- User education and MT system customization (e.g. building dictionaries), which are major factors in achieving the best possible translation results, are often easier to accomplish in smaller organizations than in larger ones.

The Bottom Line

Since its inception, machine translation has been a highly controversial technology, and it will probably continue to be so for some time. Much of this controversy is based on false assumptions about what machine translation can do and who might benefit from using this type of technology. Let me say it loud and clear: In general, the commercial machine translation systems available today cannot replace human translators, especially when those MT systems are operated by users who have no linguistic background. However, when the goal is to improve the efficiency of the human translation process or to create comprehensible translations in environments where human translation is not an option, and when these systems are operated by trained and motivated translation professionals, then machine translation is and has been a very powerful solution.

_____

1. **Systran Software Inc. 2007.** Systran Case Study: Symantec. *Systran Software Inc. website.* [Online] 2007. [Cited: June 6, 2008.] www. systransoft.com/download/case-studies/2007.12.Symantec.pdf.

2. **Microsoft Corporation. 2008.** Machine Translation - Home. *Microsoft Corporation website.* [Online] 2008. [Cited: June 6, 2008.] http://research. microsoft.com/nlp/projects/mtproj.aspx.

3. **Institute of Standards and Technology. 2006.** NIST 2006 Machine Translation Evaluation Official Results. *National Instititue of Standards and Technology website.* [Online] November 1, 2006. [Cited: June 6, 2008.] http://www.nist.gov/speech/tests/mt/2006/doc/mt06eval_official_results. html.

4. *Fully Automatic High Quality Machine Translation of Restricted Text: A Case Study.* **Muegge, Uwe. 2006.** London: The Association of Information Management (Aslib), 2006. Proceedings of the Twenty-eighth International Conference on Translating and the Computer. ISBN 978-0-85142-5.

Greg Marshall
Senior Marketing Manager,
Language Line Services

Chapter 21

Interpretation and Other Language Services

by Greg Marshall

Our world is shrinking.

More now than at any time, in business and in everyday life, we meet people of other countries, cultures and languages. But while modern technology brings us all closer, we continue to have language barriers between us. How can businesses and organizations bridge the communication gap? Using professional interpretation services is increasingly the way to address these challenges.

Reducing Language as a Barrier to Doing Business

Immigration to the United States has not slowed a bit in the current economic environment. In fact, more people than ever are coming to the US. According to the US Census Bureau, a newly arrived immigrant enters the United States every 19 seconds.

Government agencies, 911 emergency services and court systems increasingly need language services. Organizations of all types are finding that promoting their products and services in multiple languages is smart business.

Primary modes of language interpretation continue to be over-the-phone, on-site (face-to-face) and over-video.

Over-The-Phone Interpretation (OPI)

Demand for telephone interpretation has maintained an upward trajectory since Language Line Services created the industry more than 25 years ago.

During over-the-phone interpretation, interpreters manage a three-way conversation between an organization's representative and their limited-English speaking customer. Interpreters analyze each message for both content and context and then select words that most accurately convey the true meaning of what is said.

Interpretation is not word-for-word, but meaning-for-meaning. So, regardless of the language or education level of the caller, well-qualified interpreters provide accurate and thorough first-person exchanges.

Qualified interpreters are fluent in English and the language of the customer. They have an intimate familiarity with both cultures, are well

versed in the topic (health care, finance, legal, emergency service, etc.), and demonstrate superior customer service.

On-Site Interpretation Service

On-site interpretation service, on the other hand, puts Language Line Services' clients, and their limited-English speaking customers, face to face with qualified interpreters. The service is ideal for organizations that need regular interpreting or the constant attention of an interpreter with specialized training.

On-site, face-to-face interpretation is commonly used in intimate professional situations such as trials, legal depositions, doctor's appointments, worker's compensation meetings, and multilingual business meetings. It is also ideal for social events like receptions, facility site visits, and guided tours.

This type of interpreting is also used in large meetings where direct, interpersonal communication is impractical. The presenter and interpreter speak simultaneously, allowing the audience to hear the presentation in their native languages (sessions of the United Nations are examples of this type of interpretation).

Services for American Sign Language (ASL)

Oftentimes deaf or hard-of-hearing patients are left waiting for periods of time before a qualified American Sign Language (ASL) interpreter is available. Language Line® Video Interpreter Service helps minimize that wait with a video-based interpretation service for use with sign language and spoken conversations during which a visual component (monitor with speakers) is used.

To make this service possible, state-of-the-art technology is employed to enable high-quality interactive video communication with live language interpreters. All it takes is a T1 Internet connection, and the organization can use Language Line Services' equipment to access this useful video interpreting service.

Video interpreting has demonstrated significant positive results for both patients and caregivers. The addition of a visual component during interpreting has lead to more accurate diagnoses and treatments. The service is particularly valuable because instant access to specially trained medical interpreters can ensure clear communication with patients.

Video interpreting increases an organization's service efficiency by enabling the staff to quickly respond to limited-English speaking and deaf

patient needs. It helps save companies thousands of dollars each year by supplementing on-site interpreters. And it can protect organizations from potential liability claims filed under the American with Disabilities Act or Title VI of the Office for Civil Rights.

Secure Interpretation Over the Phone

Organizational staff often find themselves face to face with customers who speak little English. When they call for an interpreter, they are forced to pass the phone back and forth. This can often run the risk of missing vital information. Conversely, if a speakerphone is used, sensitive information can unintentionally be revealed to anyone within earshot.

The Language Line® Phone provides private, three-party conversations with interpreters with just the push of a single button. This specialized phone features two handsets and allows employees and customers, standing side by side, to communicate with the interpreter.

The phone promotes easy conversation flow, with no awkward midstream pauses during the conversation. A pre-programmed number allows one-touch access. Privacy is maintained because no speakerphone is necessary, though it is built in. And no special user training is needed, for employees or customers.

Phone Based Education

Language Line® University, the testing, training and certification (medical and court) division of Language Line Services, prepares bilingual staff for the demanding challenges they face as professional interpreters. It can train and test staff members to ensure customers receive professional in-language service and support. Testing can also be done to assess the proficiency of an organization's bilingual staff, aiding in the screening process for new employees.

All training and testing takes place over the phone, so sessions can be scheduled at the organization's convenience. Language Line University is lead by a dedicated and highly qualified faculty. Its training curriculum and testing services—which include language proficiency, English proficiency, interpreter skills, medical certification, and court certification tests, to name just a few—remain unequaled in the industry. National industry experts recognize phone-based education as a highly effective tool in testing and assessing interpreter skills.

Language Line University, in partnership with the International Medical Interpreters Association (IMIA), is at the forefront of establishing the first-ever national medical interpreter certification.

Direct Response Services

Forty million limited-English speaking people live in the United States today. Multicultural marketing is the fastest growing segment of the advertising industry. And companies across the nation are recognizing the value of promoting their products and services in the languages customers understand—their own.

Language Line® Direct Response makes it possible for organizations to have interpreters greet their caller first in their own language, identify their needs, and then introduce the organization's representative to the limited-English speaking caller.

When customers call, they hear an automated message in their language and are then connected to an interpreter. The interpreter greets the customer and then, acting on behalf of the organization, connects the customer to the appropriate service representative and remains on the line to assist.

LanguageTrak

Language Line® LanguageTrak is the first product of its kind in the language services industry. It tracks demographic information, in real-time, on emerging trends that affect native speakers of more than 170 languages nationwide. The service is designed to help businesses anticipate language trends and other cultural demographic changes occurring in the communities they serve. It alerts subscribers to changing geographic language use patterns so they can quickly and effectively adapt to evolving customer needs, providing a competitive advantage many organizations need.

Global Language Solutions Make a Difference

Language Line Services has been the leader in over-the-phone interpreting for more than 25 years, operating in several countries with a staff of interpreters 5,000 strong. The company created the over-the-phone interpretation industry, and established its first interpreter education, training, and certification programs. It pioneered the study of language use patterns for business forecasting applications and continues to lead the way with services that are available around the clock, around the world.

Thousands of times a day, Language Line Services' professional interpreters, and those who support them, help customers, patients and residents communicate their needs in the languages they prefer … their own.

David José
Business Development &
Marketing Director,
Lingo Systems UK

I have lived in the UK
and the US working in
a variety of sales and
marketing roles; I noticed
when people saw my
written name in the US,
they referred to me as
Hosé David and asked
whether I was from
Central America. I am now
back in my hometown of
London, and safely write
my Portuguese-roots name
as David José. This has
taught me that knowing
your country and language
can be important, which
also makes working with
my colleagues from all over
the world at Lingo Systems
such a reward.

Chapter 22

Localisation and Translation from a European Perspective

Located in Canary Wharf, London, Lingo Systems' European headquarters is at the heart of a multilingual and multicultural continent. Communication occurs in every possible way from email to fax, to Word documents, PDFs and advertisements. Document translation is often the first step towards full localisation. While the steps in the process may vary, the goal—quality translation—is the same.

Quality is paramount and there are things you can do to ensure your translation projects are a success. These apply to a few sentences in an email, a fifty-page operating procedure, or a mix of PowerPoint, PDF, Word and Artwork documents for a multi-country campaign into 15 languages.

The following pointers are designed to help you communicate with your translation provider, thereby ensuring accurate estimates, higher quality and faster turnaround times.

Start with the source text: It will save time and money

Problems in the original source text are often the root cause of problems that arise during the translation process. Resolving these problems early before

> TIP: A professional translator can translate 2,000-2,500 words per day, so if your document is 15,000 words long, it can take one translator eight days to translate. Additional time is required for proofreading and this can impact your turnaround times.

they are sent to your provider will remove the risk of schedule impacts and additional costs. Localisation costs have a direct relationship with word counts. It therefore makes business sense to review your source for duplication. It can be wise to look for opportunities to shorten sentences, simplify vocabulary and punctuation and remove or revise ambiguous images. Achieving the most efficient and highest quality translations is a result of planning for localisation throughout the entire business process.

Clearly state the final use of the documents

Communicating the final use of your documents to your translation provider will ensure you only pay for the service you need. Are you translating an email from Portuguese to English for personal use or a marketing brochure from English to German to use in sales presentations?

Each project will require a different level of service commonly referred to as "for information" or "for distribution." Depending on your needs the cost and the turnaround time to complete the translation can vary. Typically, a document required for information purposes will be translated by a native-speaking translator of the target language but not proofread. Documents intended for distribution purposes will be translated, proofread and quality checked to ensure the document reads as if it were originally written in the target language. In the same way you would proofread a document written in English, the proofreading stage allows a second native-speaking translator to refine the language, make better word choices and convey the original message of the English document more accurately. Lastly, the quality check ensures all text was translated and formatted according to the project manager's instructions.

> TIP: To create a rough estimate of words based on page count, use 250 words per page. This will increase for text-heavy legal documents and decrease for documents with graphics.

## Know your target audience: The British play football, not soccer!

Part of knowing your target audience is to know which language they speak—a factor that, ironically, can get overlooked. Naturally, we translate into Spanish, Portuguese, Chinese, German and English, but knowing the target audience will help determine if you need Spanish for Spain or Mexico, Portuguese for Portugal or Brasil, Traditional or Simplified Chinese, German for Germany, Switzerland or Austria and even English for UK or US. If you don't know, discuss it with your provider before starting the project. Translating for the wrong country or region could mean your message is not communicated in a culturally appropriate manner.

Alex Macnab
Business Development Manager,
Lingo Systems UK

Having worked in financial services in the US, I never imagined I would end up working in the translation industry! It has been a great adventure getting to know London and the many interesting people who work in the localization business. I was born in London, moved to Dallas, Texas when I was 4 and moved back to London in 2006. Europe has provided inspiration for one of my favorite hobbies, photography. (It's not too bad for golf and rugby, either!)

UK Translation Process
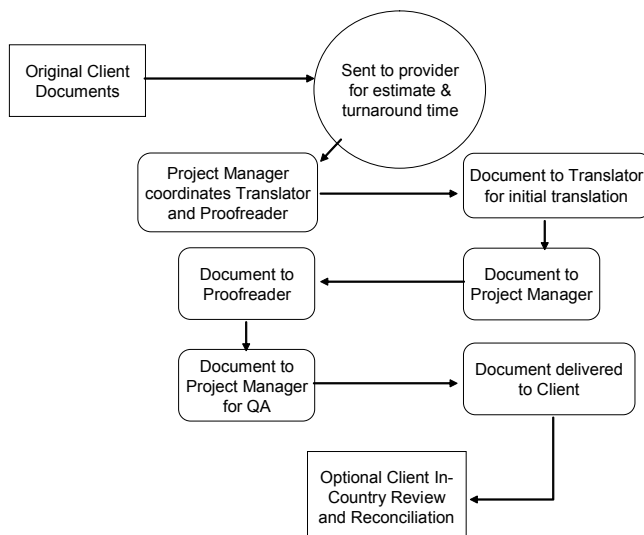
Silvia Sanchez
Translations Manager,
Lingo Systems UK

The masterpiece of two Spanish artists, their creative flair hasn't skipped a generation. The definitive tomboy, my nature means you'll regularly find me behind the lens, the bar or the DJ decks. But put yourself in safe hands, as it's on the desk at Lingo Systems in the London office that you'll find my best work.

Provide training and supplemental reference materials

Translators love cheating! Well not really, but they do prefer to work with reference material. As language professionals, a translator's goal is to translate your text and message accurately into a specific language. Providing translators with previously translated material, with which you are satisfied, will help them create translations with the terminology, style and feel that you want. Reference material could be existing marketing brochures, human resource documents, legal documents, even websites that convey your company's preferred communication style. Your company may already have an English style and terminology guide that can be leveraged. It is often said that translation is an art, not a science. There may be multiple words or phrases to express the same meaning and translators continually decide how to express these ideas with your style and terminology in mind. Reference material will make those decisions easier and ultimately will create a translation with which you and your colleagues will be satisfied.

Please provide feedback: It will save you money

One definition of feedback is "the process in which part of the output of a system is returned to its input in order to regulate its further output." Now that's a mouthful! Simply put, your translation provider wants to hear from you and your colleagues

> TIP: Free online translators like "Babel Fish" or "Google Translate" are called machine translation. While useful in some instances, they should never be used to completely replace professional translation and can produce some humorous results!

with any feedback. This enables them either to amend the translation or to record the feedback to improve future translations. Most feedback is done by a colleague who speaks the language or an in-country reviewer. They may edit or refine the translation to fit their personal style or the specific terminology of the company. These changes can provide valuable feedback and reference material to translators on future translations.

ASAP is not a deadline

Turnaround times and deadlines are often the most important consideration when choosing a provider for a project or ongoing work. Most providers offer "normal" and "rush" turnaround times and quality work can be achieved under both circumstances. Ideally your provider will have the recommended time to translate and/or proofread your documents. Any project that is rushed requires additional resources and attention to detail and the translation industry is no exception. If you have a project that

requires translation but have not finished the original documents yet, let your provider know what is coming. This allows your provider to line up resources, potentially saving you money and time!

Client and information confidentiality

Client confidentiality and the vigilant handling of sensitive materials are critical in our industry. We are often entrusted with legal, financial, human resource and marketing documents before employees, lawyers or the general public have any knowledge of them. In recent times, high profile

TIP: Languages can expand by up to 30% when translated, so it is important to consider this when thinking about page layout and white space in creating a document destined for translation.

failures in data protection have highlighted just how catastrophic a lapse in protection can be. Every year we translate millions of documents and, simply put, if we failed to protect confidential information we would go out of business. We implement rigorous processes across our entire business and carry the highest level of insurance (in the language industry) for third party information protection.

oops

On a menu in Beijing during the 2008 Olympics:

Deep Fried & Look Like Squirrel

Braised Dork with barboo shoot

A sign from Peru:

"friend proves our variety of pizzas

prepared in firewood furnace … enters

to the bottom …"

Glossary

.BMP (BMP)

A standard bit-mapped graphics format used in Windows. Files end with .BMP extension.

.GIF (GIF)

Graphics Interchange Format. A bit-mapped graphics file format used by the World Wide Web. It features lossless data compression and is best for computer-generated, nonphotographic images. Files end with .GIF extension.

(See also Lossless.)

.h and .hpp files

Header files. These are files used in programming (typically C/C++) to identify and define common items used throughout the program.

.JPG (JPEG)

Joint Photographic Experts Group. A lossy compression-type graphics format for color files. Can compress files to 5% of their original size with (some) loss of picture quality. Best for photographic images. Files end with .JPG extension.

(See also Lossy.)

.NET

Microsoft platform for applications that work over the Internet.

.PCX (PCX)

A graphics file format used by PC graphics applications. This widely used file format employs lossless compression. Files end with .PCX extension.

(See also Lossless.)

.PDF (PDF)

Portable Document Format. A file format created by Adobe Acrobat, primarily for read-only use with Acrobat Reader. Can be edited with the full version of Acrobat. PDF files capture formatting and layout data from files created with another application, allowing others without that source application to view properly formatted documents via Acrobat Reader on any system supported by Acrobat Reader. Files end with .PDF extension.

.SHG files

Bitmaps with a hotspot overlay.

(See also Hotspot.)

.TIF (TIFF)

Tagged Image File Format. Widely used file format for storing bit-mapped images on both PC and Macintosh platforms. Commonly used for scanned images. Files end with .TIF extension.

Active Server Page (ASP)

An HTML page where one or more scripts are processed by an MS server prior to the page being displayed to the user.

ActiveX

A Microsoft program development technology that allows data to be shared among different applications. Conceptually similar to Java, ActiveX has a significant presence in web-based applications.

A-Link

A linking macro provided in WinHelp that allows jumps based on keywords rather than specific context strings. A-links do not have to be localized. A-links are never seen by the end-user and are used only by the help system.

American Sign Language (ASL)

The dominant sign language of the Deaf community in the United States, in the English-speaking parts of Canada and in parts of Mexico. Although the United Kingdom and the United States share English as a spoken and written language, British Sign Language (BSL) is quite different from ASL and not mutually intelligible.

AMT

Automated Machine Translation (AMT) and Caterpillar Technical English (CTE) are development project collaborations between Caterpillar, Inc., and Carnegie Mellon University to further improve the creation and translation of technical documentation into three core languages: Spanish, French and German.

ASCII

American Standard Code for Information Interchange. A standard for assigning numerical values to the set of letters in the Roman alphabet and typographic characters.

Automatic Content Enrichment (ACE)

A bridge between single language websites and localization, ACE technology associates English words and phrases on web pages with pop-ups containing information in a user's native language.

Back translation

The process of translating a document that has already been translated into another language back to the original language - preferably by an independent translator.

Bidirectional

A writing system in which text is generally flush right, and most characters are written from right to left, but some text is written left to right as well. Arabic and Hebrew are the only bidirectional writing systems in current use.

Bidirectional text (bidi)

A mixture of characters within a text where some are read from left to right and others from right to left. *Bidirectional* or *bidi* refers to an application which allows for this variance.

BinHex

Binary hexidecimal. A widely-used encoding scheme that converts binary data into ASCII characters. BinHex encoding is especially common on MAC platforms. Files end with .HQX extension.

Bitext

A merged document comprised of both source-language and target-language versions of a given text. Bitexts are generated by a piece of software called an alignment tool, which automatically aligns the original and translated versions of the same text.

Bitmap

A graphic for which the color of each pixel is defined by one or more bits (1 bit for black/white, 4 bits for 16 colors, 8 bits for 256 colors, etc.).

Blog

Shortened from "weblog," this is a web application which contains periodic time-stamped posts on a common webpage. Blogs range from individual diaries to arms of political campaigns, media programs and corporations, and from having one occasional "blogger" (author) to having large communities of writers.

Bloggerati (sing. bloggerato)

Adapted from *literati*, the term refers to the "A-list bloggers"—popular and/or celebrity bloggers in the blogging community.

BRIC

Term used to refer to the combination of Brazil, Russia, India and China.

Cascading Style Sheets (CSS)

A way of implementing styles in HTML or XML. By combining the styles from several sheets, or using specific rules to override general rules, you can "cascade" the information across multiple pages.

Call out

A small text box referring to an element or feature in a graphic.

CAT

Computer Aided Translation. A broad term used to describe computer applications that automate and assist with the act of translating text from

one language to another. CAT tools are highly effective in improving translation productivity and quality (e.g., Trados Workbench and associated utilities).

CE marking

The letters *CE* are the abbreviation of the French phrase *Conformité Européenne* that literally means European conformity. CE marking on a product is a manufacturer's declaration that the product complies with the essential requirements of the relevant European health, safety and environmental protection legislations.

Central European fonts

Specific fonts used for displaying Central and Eastern European languages.

Character

A symbol standing for the smallest abstract component of a writing system or script, including sounds, syllables, notions or elements, as opposed to glyphs.

(See also Glyph.)

Chunk

Depending on the level in which it is stored, this term is used to describe a subset of content stored in a Content Management System (CMS). A Chunk can be a word, phrase, sentence or paragraph. Chunks are combined by the CMS to create a document.

(See also Content management system.)

CJKV

The abbreviation for the languages Chinese, Japanese, Korean and Vietnamese.

CMYK

Cyan Magenta Yellow Black. A color model in which all colors are described as a mixture of these four process colors. CMYK is the standard color model used in offset printing for full-color documents. Also called four-color printing.

(See also RGB.)

CNS

The Chinese National Standard (CNS) 11643-1992 defines a total of 48,027 characters and applies the EUC-TW (extended UNIX code-Taiwan) to one-, two- and four-byte encoding.

Compiling

Converting a program written in a high-level programming language from source code into object code. Source code must be compiled before it becomes an executable program.

(See also Decompiling.)

Computational linguistics

The engineering of systems that process or analyze written or spoken natural language. It is concerned with the computational aspects of the human language. Its goal is to provide computers with the ability to produce and interpret human language.

Computer-based training (CBT)

A form of education in which the student learns by executing special training programs on a computer.

Computer code

The computer readable code that makes up a program. Also called object code or machine language.

(See also Executable.)

Concatenation

A programming method used to avoid creating a number of repetitive messages by starting with a base sentence containing variables that grab the desired elements when the software is run.

Consecutive interpreting

The interpreter begins his or her interpretation of a complete message after the speaker has stopped producing the source utterance. At the time that the interpretation is rendered, the interpreter is the only person in the communication environment who is producing a message. Normally, in consecutive interpreting, the interpreter is alongside the speaker, listening

and taking notes as the speech progresses. When the speaker has finished or comes to a pause, the interpreter reproduces the message in the target language, in its entirety and as though he or she were making the original speech.

Content management system (CMS)

Tools that automate the process of storing, creating, maintaining, publishing, and updating content so that users can subsequently find and retrieve large amounts of data. There are many different CMSs available on the market but many of them are not true CMSs but rather file management systems. A true CMS allows authors to place their content in a centralized system that is shared across the organization. This content can then be reused amongst authors and the content can be published to many different formats including Documentation, Help, Web, etc.

(See also Globalization management system.)

Controlled languages

Subsets of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity. Also, stylistic rules—such as not using certain verb tenses or the passive voice—can be created, depending upon the group or organization and its language usage goals.

Controlled vocabulary

The standardization of words which may be used to search an index, abstract or information database. There is usually a published listing or thesaurus of preferred terms identifying the system's vocabulary.

Corpus (plural corpora)

A large body of natural language text used for accumulating statistics on natural language text. Corpora often include extra information such as a tag for each word indicating its part-of-speech and perhaps the parse tree for each sentence. Also, a large body of source-language text used for a translation.

Cropping

Trimming the edges of a graphic to make it fit or to remove unwanted sections.

Crowdsourcing

A neologism for the act of taking a task traditionally performed by an employee or contractor, and outsourcing it to an undefined, generally large group of people, in the form of an open call. For example, the public may be invited to develop a new technology, carry out a design task (also known as *community-based design* and distributed participatory design), refine an algorithm or help capture, systematize or analyze large amounts of data. The translation community has several projects underway that utilize thousands of people contributing to translation memories that are available online.

Cultural assessment

Analyzing an individual's cultural preferences through comparative analyses. Allows individuals to acquire the awareness and knowledge necessary for building effective skills and behavioral adaptations for multicultural management and business.

Cultural orientation

Developing cultural self-awareness and effective behavioral strategies to minimize the cultural gaps that occur when contrasting value orientations of different social groups.

Database

An organized collection of data managed by a program that allows users to add, delete, and change the data.

(See also DBMS.)

Data mining

Analysis of data in a database using tools which look for trends or anomalies without knowledge of the meaning of the data. Data mining uses computational techniques from statistics and pattern recognition.

DBE

Double-Byte Enabling. Re-engineering original source code to support the input, display, and manipulation of double-byte language character sets.

DBMS

Database Management System. This program enables users to manage and utilize a database. It is the interface that facilitates a multi-user system; it also tracks where data is stored in the storage media so that other programs do not have to duplicate this information.

Desktop publishing (DTP)

Using computers to lay out text and graphics for printing in magazines, newsletters, brochures and so on. A good DTP system provides precise control over templates, styles, fonts, sizes, color, paragraph formatting, images and fitting text into irregular shapes.

Dialect

A variety of a language used by people from a particular geographic area. The number of speakers and the area itself can be of arbitrary size. A dialect is a complete system of verbal communication—oral or signed but not necessarily written—with its own vocabulary and/or grammar.

DITA (Darwin Information Typing Architecture)

An XML-based architecture for authoring, producing and delivering technical information. This architecture consists of a set of design principles for creating "information-typed" modules at a topic level and for using that content in delivery modes such as online help and product support portals on the Web.

Decompiling

Opposite of compiling. Changing an application from computer code back into source code. Sometimes referred to as reverse engineering.

(See also Compiling.)

Dialog boxes

The rectangular windows used by a program to display information or request information in a User Interface (UI) (Windows or Mac).

DLL

Dynamic Link Library. A file that contains executable functions or data for applications. Several DLLs come with Windows and are used by many applications, others are written for specific applications. Files end with .DLL extension.

Dots Per Inch (DPI)

A common measurement of resolution used in printing to describe the density of an image or character. Refers to the number of dots of ink a printer is able to print per square inch. In general, the higher the DPI, the higher the quality of the printed image.

Double-byte character

A character defined with two bytes (16 bits) instead of one byte (8 bits).

Double-byte enabled

A program that can handle double-byte languages.

(See also DBE.)

Double-byte languages

Languages such as Chinese, Japanese and Korean (CJK) that use twice as much memory because their characters are more complex and graphical than Roman alphabet letters, which require only one byte of information for each letter. CJK languages are character-based with each character referring to an idea as opposed to a specific shape of the character or an object.

Drivers

Specialized programs that allow communication between peripherals (printers, scanners, video cards, etc.) and the computer.

Dubbing

In filmmaking, the process of recording or replacing voices for a motion picture. The term is most commonly used in reference to voices recorded that do not belong to the original actors and speak in a different language than the actor is speaking.

Dynamic content

Data or content of a website that is stored in a database and is supplied to the user on the fly, based on what is requested (usually through a form) by the user.

Embedded graphic

A graphic is known as an embedded graphic if all the information for it is stored in a document and not in a separate file.

(See also Referenced graphic.)

Encoding scheme

Rules for assigning numeric value (code points) to characters. Encoding is a method by which a character set is turned into computerized form for transmission and preservation.

European Union (EU)

An intergovernmental and supranational union of 27 democratic member states. The EU was established under that name in 1992 by the Treaty on European Union (the Maastricht Treaty).

Exact match

A term used to describe matching within a translation memory. Also known as 100% match and Repetitions. This is a segment stored in the translation memory, identical to the source segment that is to be translated. At Lingo Systems, this text is marked green in the pre-translated file and is charged at 25% of the full word rate.

Executable

A program that can be run (executed) on a computer.

Extended UNIX code (EUC)

A multibyte encoding design used to encode Japanese, Chinese, Korean and Taiwanese on UNIX systems.

FIGS

Abbreviation for the commonly used language set of French, Italian, German, and Spanish.

File transfer protocol (FTP)

An alternative to e-mail for transferring files utilizing the Internet. It is often faster and more reliable to use FTP software for large file transfers. It can be accessed using an Internet browser or FTP client software.

Full match

A source text segment which corresponds exactly (100%) with a previously stored sentence in a translation memory (TM) tool.

Functional QA

Testing of a software application or program to ensure that the localization process does not affect the functionality of the software and that the content displays correctly on the screen.

Fuzzy logic

A logic that allows the concept of partial truth-truth values between "completely true" and "completely false." Used to create near matches instead of exact matches during searches, and in artificial intelligence programs.

Fuzzy match

A term used to describe matching within a translation memory. A segment that is similar (but not identical) to the sentence or phrase the translator is currently translating. At Lingo Systems, this text is marked dark yellow in the pre-translated file and is charged at 50% of the full word rate.

GIIN

See Globalization.

Gist translation

A less-than-perfect translation performed by machine or automatic translation.

Globalization (g11n)

The process of conceptualizing your product line for the global marketplace so that it can be sold anywhere in the world with only minor revision. Addresses the business issues associated with launching a product globally, such as integrating localization throughout a company after proper internationalization and product design. In g11n, the common abbreviation for globalization, *11* refers to the eleven letters between the g and the n.

Globalization management system (GMS)

A system that focuses on managing the translation and localization cycles and synchronizing those with source content management. Provides the capability of centralizing linguistic assets in the form of translation databases, leveraging glossaries, and branding standards across global content.

(See also Content management system.)

Glossary

In the context of localization, a glossary is a list of source language terms paired with a list of corresponding terms in the target language.

(See also Terminology list.)

Glyph

The shape representation or pictograph of a character.

(See also Character.)

GUI

Graphical User Interface. The part of a software application that is visible to the end-user.

Hiragana

A flowing phonetic subscript of the native Japanese writing system. In Hiragana, all of the sounds of the Japanese language are represented by 50 syllables.

Hotspot

The part of a graphic in a hypertext document that jumps to another location when the user clicks on it. Similar to a hypertext link.

HTML

Hyper Text Markup Language. A markup language that uses tags to structure text into headings, paragraphs, lists and links, and tells a web browser how to display text and images on a web page.

(See also SGML and XML.)

HTML Help

A Microsoft Help system format based on HTML.

I18N

See Internationalization.

Ideographic language

A written language in which each character represents an idea, concept or other component of meaning, rather than pronunciation alone. Japanese Kanji, Chinese Hanzi and Korean Hanja are examples of ideographic writing systems.

Integration testing (Interoperability)

Confirmation that two or more systems (computers, communication devices, networks, software, and other information technology components) are able to interact with one another and exchange data according to a prescribed method in order to achieve predictable results. (ISO ITC-215.)

International brand development

The process of giving a product the look and feel of having been developed in the target country. Requires that your brand or message be clear, easily identifiable, and culturally acceptable to the target market. A successful brand can be conveyed clearly and concisely, regardless of the language or writing source.

Glossary

Internationalization (i18n)

The process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for redesign. In i18n, the common abbreviation for internationalization, the 18 refers to the eighteen letters between the i and the n.

Interpretation

Translation of spoken words from one language to another. Simultaneous interpretation occurs when the interpreter translates the speech as it is being given. The more traditional interpretation practice is called consecutive interpretation, wherein an entire thought is expressed by the speaker, the speaker pauses, and the interpreter converts the content for the target language speakers to hear. The skill set of an interpreter is different than that of a translator.

ISO

International Organization for Standardization. A network of national standards institutes from 145 countries working in partnership with international organizations, governments, industry, business and consumer representatives. ISO acts as a bridge between public and private sectors.

Java

A platform-independent, object-oriented programming language. Java can add animation, spreadsheets, and information processing features that HTML cannot provide.

JavaScript

An open-source scripting language for design of interactive websites. JavaScript can interact with HTML source code, enabling web developers to use dynamic content. For example, JavaScript makes it easy to respond to user-initiated events (such as form input) without having to use common gateway interface (CGI).

Kana

The two Japanese syllabaries—hiragana and katakana.

Kanji

The Chinese characters that are used in the modern Japanese logographic writing system along with hiragana, katakana and the Hindu-Arabic numerals. The Japanese term *kanji* literally means Han characters. Despite the existence of some 13,000 kanji characters, these alone do not suffice to write Japanese. Hiragana characters are also required to express grammatical inflections.

Katakana

A Japanese syllabary, one component of the Japanese writing system along with hiragana, kanji and in some cases the Latin alphabet. The word *katakana* means *fragmentary kana*, as they are derived from components of more complex kanji. Katakana are characterized by short straight strokes and angular corners and are the simplest of the Japanese scripts. Katakana and hiragana both render the same syllables, but katakana is angular and used largely to spell words borrowed from other languages, while hiragana is cursive and is used more frequently to spell native Japanese words.

Kerning

The space between two text characters.

(See also Tracking.)

K-link

A linking macro provided in WinHelp that allows jumps based on keywords rather than specific context strings. K-links require translation.

L10N

See Localization.

Leading

The space between two or more lines of text. Sometimes called line spacing.

Leverage

Building current translation projects on those previously completed. Reduces the need to retranslate words and phrases previously translated. The process of using one translation for repeated sections of text.

Lexicography

The act of compiling dictionaries.

Lingua franca

A language that is adopted as a common language between speakers whose native languages are different.

Linguist

Someone who is proficient in several languages. A student or practitioner of the subject of linguistics (the scientific study of languages and their structures).

Linux

A free open-source UNIX-type operating system which runs on a number of hardware platforms.

Localization (l10n)

Adapting a software, document, or website product to various markets or localities so that it seems natural to that particular region. This may require a variety of steps including translating user interface text, modifying formats for numbers and dates, and replacing culturally inappropriate graphics or system design. In l10n, the common abbreviation for localization, the 10 refers to the ten letters between the l and the n.

Localization engineering

The process of using specific localization applications, compilers, and tools to prepare software for release in other markets or localities.

Lossless

A term used to describe compression techniques that don't lose any data. Lossless compression techniques usually reduce the size of the compressed file up to 50% of the original file.

Lossy

A term used to describe compression techniques that lose some data or details. Commonly used with graphics and video. Lossy compression techniques can compress files to around 5% of their original size with some loss of data.

Machine translation (MT)

A technology that translates text from one human language to another, using terminology glossaries and advanced grammatical, syntactic, and semantic analysis techniques.

Machine-aided translation (MAT)

Computer technology applications which assist in the translation of text from one spoken language to another, based on the concept of translation memory (TM) and the reuse of previously translated terms and sentences.

Multilingual

Refers to software that supports more than one language simultaneously, thereby allowing the end user to select multiple languages and formats. This software allows data containing multiple languages to be entered, processed, presented and transmitted multinationally.

Multilingual print production

Producing packaging, advertising, and related collateral in multiple languages for simultaneous release.

Multilingual workflow system (MWS)

A computer program which creates an environment that supports and orchestrates a range of activities that facilitate the development of multilingual products. An MWS should contain a globalization management system (GMS) for managing multilingual content, along with translation memory (TM) and machine translation (MT).

Multimedia

In computing, multimedia describes a number of diverse technologies that allow visual and audio media to be combined. Entertainment, education and advertising applications, among others, use a computer to present and combine text, graphics, video, animation and sound.

Multiple Language Vendor (MLV)

A company that offers translations into multiple languages.

Multiterm

An application made by TRADOS that indicates previously translated terminology to translators.

Nearshoring

A form of outsourcing in which an activity— for example, business processes or software development—is relocated to locations which are, generally, cheaper and yet geographically nearer than offshore locations.

New text

A term used to describe matching within a translation memory. Text where the source segments being translated do not correspond to any of the target segments. At Lingo Systems, this text is marked blue in the pre-translated file and is charged at the full word rate.

OEM

Original equipment manufacturers (OEM) buy computers in bulk and customize them for a particular application. OEMs then sell the customized computers under their own names. Therefore, OEMs are really the customizers and not the original manufacturers of the equipment.

Offshore outsourcing (offshoring)

The practice of engaging a third-party provider in another country (often on another continent or "shore") to perform tasks or services often performed in-house.

Online content

Any written content that is intended for publishing via a network or the Internet. While users can print the material, the format is often not optimized for specific layout specifications.

Online user interface

A software product or service that is delivered to the user via a network or the Internet. This software does not reside or run on a "local" computer.

Open-source software

Any computer software distributed under a license that allows users to change and/or share the software freely. End users have the right to modify and redistribute the software, as well as the right to package and sell the software.

Outsource

To hire a third-party provider to perform tasks or services often performed in-house. The third-party provider is then referred to as the outsourcer.

Parser

A computer program that takes a set of sentences as input and identifies the structure of the sentences according to a given grammar. The term *parser* is sometimes used generically in cases where the sentences are made up of information units of any kind.

Phonology

The part of linguistics that deals with systems of sounds especially in a particular language.

Pixel

Picture Element. One dot on a computer screen. The smallest image-forming unit on a display screen.

Pretranslation

Involves the preparation of files for translation where the existing files already contain related segments of previously translated data. Only 100% matches are replaced, with the result being a set of files containing both source and target language terminology.

Project management (PM)

Project management (PM) is the systematic planning, organizing and controlling of allocated resources to accomplish project cost, time and performance objectives. PM is normally reserved for focused, nonrepetitive, time-limited activities with some degree of risk.

Pseudo-localization

Translates the code strings of a product into "pseudostrings." The resulting "pseudo-language" is designed to test the impact that different aspects of localization have on the product's functionality and appearance.

Pseudo-translation

Similar to a test run which seeks to copy the translation process rather than actually produce a translation. A text string is taken and put through a translation-like process which alters it and produces a new string. The text string is frequently changed as a result of this process, so pseudo-translation is done to illustrate the potential problems that may occur when the translation is actually done.

Quality assurance (QA)

The process of assuring that the localized content (whether it is a document or a User Interface) resembles the source content as closely as possible. The process can include verification of layout and graphics to confirm a document is complete, and the testing of an application's functionalities, etc.

Referenced graphic

A graphic that appears in a document in which the information for the graphic is stored in a separate file; minimal information about the graphic is stored in the actual document.

(See also Embedded graphic.)

Resizing

During localization, software strings may expand and no longer fit within the dimensions of the source dialogs, buttons, menus, etc. When this occurs, engineers use specialized software to resize the UI elements so that the text fits within the allocated area.

Resource files

Source files that contain information to be compiled into the program. They contain the parts of the application that is seen by the user. Typical file types include: .rc, .res, .resx, .bmp, .ico, .cur.

RESX

A resource file used within the Microsoft .NET Framework to specify strings and objects in an XML format.

RGB

Red Green Blue. Blending these three colors allows computer monitors to display color images.

(See also CMYK.)

Right-to-left languages

Languages such as Hebrew, Arabic, Urdu and Farsi are written primarily right-to-left. This text flow presents significant text and graphic layout implications.

RoboHelp

An application made by eHelp. RoboHelp assists in writing help files using Microsoft Word.

RTF

Rich Text Format. A type of document that encodes formatting as text-based tags. Can be opened as text to view the tags or converted to look like a Word document (without the tags visible). Used as a source file for WinHelp.

Scaling

Changing the size of a graphic so that no distortion occurs.

(See also Sizing.)

Screen shots

A graphic image of what is seen on the computer screen. Often used in user's manuals to show how an application looks on the screen. Also called "screencaps," "screen captures," or "SCAPS".

SEA

Self Extracting Archive. A file that decompresses itself. Used on a Mac OS.

Search engine

A program designed to help find information stored on a computer system such as the World Wide Web or a personal computer. A search engine allows a user to ask for content meeting specific criteria - typically those containing a given word, phrase or name - and retrieves a list of references that match those criteria.

Search engine optimization (SEO)

A set of methods aimed at improving the ranking of a website in search engine listings. SEO is primarily concerned with advancing the goals of a website by improving the number and position of its organic search results for a wide variety of relevant keywords.

Segment

The basic unit of source text, as identified by a translation tool, that can be aligned with a corresponding translation from the translation memory. A segment is commonly defined as the content from one paragraph break to the next, usually a sentence, but a segment can be a header, items in a list, cells in a table, a paragraph, etc.

(See also Translation memory.)

Semantic Web

An extension of the World Wide Web that provides a common framework allowing data to be shared and reused across application, enterprise and community boundaries. It is based on Resource Description Framework (RDF), which integrates a variety of applications using XML for syntax and URLs for naming.

SGML

Standard Generalized Markup Language. SGML is an ISO standard for marking text files to show how they should be formatted, defining three document layers: structure, content and style. HTML is a specialized application of SGML rules.

(See also XML and HTML.)

Sight translation

With sight translation, the input is visual (the written word) rather than oral (the spoken word). Reading comprehension is an important element of sight translation.

Simplified Chinese

A Chinese character set used in mainland China and Singapore, modified to be written with fewer strokes per character.

(See also Traditional Chinese.)

Simship

A term used to refer to the simultaneous shipment of software products in different languages or with other distinguishing differences in design.

Simultaneous interpreting

Occurs while the source speaker is speaking, as quickly as the interpreter can reformulate the message into the target language.

Single language provider (SLP)

A company that specializes in translating into only one language.

Single sourcing

A method of using one source of stored content to generate multiple types of documents on multiple platforms.

(See also Content management system.)

Sizing

Changing the dimensions of an image or graphic. Sizing can cause distortion or loss of image quality.

(See also Scaling.)

Source code

The human-readable code that is compiled to make a program. Some types of source code are C++, Java, and Visual Basic.

(See also Source file.)

Source file

A file containing the source material that is used to create the translated product during a localization project.

(See also Source code.)

Source language

A language that is to be translated into another language.

String tags

Tags used in strings to mark where something will be added. For Example: "%s" = another string, "/n" = a return character, and "/t" = a tab.

Strings

Groupings of characters (letters, numbers, and/or punctuation marks) that are used in programs such as error messages, button labels, etc. Often strings are enclosed in single or double quotes. Strings need to be translated if they contain text that will be seen by the user.

Style sheet

A document or template that defines the style and layout of a document. Contains instructions for margins, fonts, page size, spacing, etc. Aids in the consistent appearance of pages in a large document. Also known as a style guide.

Terminology list

The terminology list is created as a reference for linguists (translators), and is usually specific to a project. It provides the linguists with the English source word and the target language equivalent. Terminology lists are created by the linguists and approved by the client prior to translation. A list of terms and descriptions are recommended for each specific case.

(See also Glossary.)

Terminology management

Primarily concerned with manipulating terminological resources for specific purposes—for example, establishing repositories of terminological resources for publishing dictionaries, maintaining terminology databases, ad hoc problem solving in finding multilingual equivalences in translation work or creating new terms in technical writing. Terminology management software provides the translator a means of automatically searching a given terminology database for terms appearing in a document, either by automatically displaying terms in the translation memory software interface window or through the use of hotkeys to view the entry in the terminology database.

Terminology manager

A computer technology application tool that assists in the translation of text from one spoken language to another.

Text expansion

The increase in the total number of characters that often occurs during translation.

Text extraction

Manually or electronically pulling text out of a source file (Quark, Illustrator, PDF) and placing it into a Word file for use by a linguist.

Tracking

The average space between characters in a block of text.

(See also Kerning.)

Traditional Chinese

A Chinese character set used everywhere except in mainland China and Singapore. This set is consistent with the original Chinese ideographic form that is several thousand years old.

(See also Simplified Chinese.)

TRADOS Translator's Workbench

An application that assists a translator by showing how similarly translated sentences were translated. This software program is used to store linguist-translated text and display it when previously translated phrases appear in a word file. Helps to assure consistency and reduce redundant work.

Translation

The process of converting a source language into a target language. An understanding of the context or meaning of the source language must be established in order to convey the same message in the target language. Translation is a crucial component of localization.

Translation management system (TMS)

A category of software applications and web services that incorporates traditional translation memory functionality with workflow. Some tools on the marketplace are hosted by localization vendors, thereby blending translation services with workflow; others can be implemented as stand-alone workflow with independent translation resources.

Translation memory (TM)

A database in which previous translations and corresponding source text are stored for future use. New source text is automatically paired with these prior translations through the use of database technology. The matching process identifies new text, fuzzy matches and 100% matches and repetitions. The linguist uses the Workbench to manage this process.

(See also TRADOS Translator's Workbench.)

Translation Memory eXchange (TMX)

An open standard, based on XML, which has been designed to simplify and automate the process of converting translation memories (TMs) from one format to another.

(See also Translation memory.)

Translation memory system

A tool for computer-aided translation. The TM stores the original text and its human translation in manageable units. The TM system proposes the translation whenever the same or a similar unit occurs again.

Translation portal

A website or service that offers a broad array of resources via the Internet, thus providing a marketplace for translation agencies, freelance translators and customers to exchange services.

Translation Unit (TU)

A single segment pair of source and translated text stored in the translation memory. The translation unit may be a single word, a phrase, one or more sentences, or even a larger unit.

Unicode

The Unicode Worldwide Character Standard (Unicode) is a worldwide standard for encoding modern and classical written text in virtually all of the world's languages. Unicode identifies a unique number for every character, independent of operating system, platform, or language. This allows data from several languages to be transported across different server or computing environments without fear of corruption.

Verification testing

Confirmation of any testable requirement, including functional testing of hardware and software system components, linguistic correctness, compatibility testing of one component to another, design verification, compliance to industry and international standards, and third party interoperability. Testing is conducted on localized, native operating systems and hardware.

Voice-over

Refers to a production technique where a disembodied voice is broadcast live or pre-recorded in radio, television, film, theater and/or presentation. The voice-over may be spoken by someone who also appears onscreen.

WinHelp

Short for Windows help file. WinHelp is also the name of the application that runs Windows help files (.hlp).

Written Chinese

Written Chinese refers to the thousands of symbols or Chinese characters used to represent spoken Chinese, along with rules and conventions about how they are arranged and punctuated. Chinese characters do not constitute an alphabet or a compact syllabary. Instead, they are built up from simpler parts representing objects or abstract notions, although most characters do contain some indication of their pronunciation.

XLIFF (XML Localization Interchange File Format)

Specifically designed to support the localization of data and has features for updating strings, revision control, marking different phases of the localization process, word count calculations, the provision of alternative or suggested language translations, among others. XLIFF is an open standard.

(See also XML.)

XSL (eXtensible Stylesheet Language)

A language for expressing style sheets, controlling formatting and other output behavior.

XML

eXtensible Markup Language. This is a pared-down version of SGML that is a universally accepted format for creating and tagging documents and data for display on the Web. It was developed, and is administered, by the World Wide Web Consortium (W3C).

(See also HTML and SGML.)

Zip file

A compressed file created by the utility application PKzip or WinZip on a PC.

At the Tran Quoc Pagoda in Ha Noi, Vietnam:
No wear the short

A sign on the lion cage at a zoo in the Czech Republic:
No smoothen the lion.

Resources, Publications, and Contacts

RESOURCES	
European Union	http://europa.eu.int
International Trade Administration, Department of Commerce	1-800-USA-TRADE http://trade.gov
U.S. Bureau of the Census, Foreign Trade Statistics	www.census.gov/foreign-trade/ www/
U.S. Chamber of Commerce	1-202-659-6000 www.uschamber.org/default.htm

U.S. Department of Commerce
        For export advice contact:
    • http://export.gov/
    "Basic Guide to Exporting"
    • http://www.export.gov/basicguide
	1-202-482-2000
	www.commerce.gov

ASSOCIATIONS	
American Translators Association (ATA) 225 Reinekers Lane, Suite 590 Alexandria, VA 22314 USA	1-703-683-6100 www.atanet.org
American Society for Testing and Materials (ASTM) 100 Barr Harbor Drive West Conshohocken, PA 19428 USA	1-610-832-9585 www.astm.org
Localisation Industry Standards Association (LISA) Domaine en Praël CH-1323 Romainmôtier, Switzerland	+41 24 453 2310 www.lisa.org
Society for Technical Communication (STC) 901 North Stuart St., Suite 904 Arlington, VA 22203 USA	1-703-522-4114 www.stc.org
Software & Information Industry Association (SIIA) 1090 Vermont Ave., NW, Sixth Floor Washington, DC 20005 USA	1-202-289-7442 www.siia.net

PUBLICATIONS	
ATA Chronicle • Publication of the American Translators Association	1-703-683-6100 http://www.atanet.org/ publications/chronicle.php
Intercom • The magazine of the Society for Technical Communication	1-703-522-4114 http://www.stc.org/intercom/
J@pan Inc • Website and magazine with useful information about doing business in Japan.	www.japaninc.com
The Localization Institute, Inc. 7601 Ganser Way Madison, WI 53719 USA	1-608-826-5001 www.localizationinstitute.com
MultiLingual, Inc. 319 North First Street Sandpoint, ID 83864 USA • **MultiLingual Magazine** Publication of record for the Translation and Localization industries. • **MultiLingual News** Free biweekly electronic newsletter featuring the latest news for the language industry.	1-208-263-8178 www.multilingual.com info@multilingual.com
Software Business 7355 E. Orchard Road, Suite 100 Greenwood Village, CO 80111 USA	1-720-528-3770 www.softwarebusinessonline.com
Technical Communication • Journal of the Society for Technical Communication	1-703-522-4114 www.stc.org/pubs/ techcommGeneral01.asp

Supercharge your Skills to Succeed in the Dynamic Global Economy

www.globalizationexecutive.com

Today's economy is characterized by increased connectivity of global markets, resource mobility, shifts in global production, unfettered global competition and turbo charging of emerging economies like China, India and Brazil. To thrive in this volatile and dynamic global economy you need high-end knowledge skills that are valuable, rare and hard to imitate. The Executive Certificate in Web Globalization Management offered by the John Cook School of Business at Saint Louis University is an innovative academic program that equips you with such high-end knowledge skills. So see how the Executive Certificate in Web Globalization can help you take advantage of emerging global megatrends:

- Megatrend 1: *The Web will be critical to achieving global expansion*. This is one of the only programs that prepares you to effectively understand the dynamics of conducting global e-business.

- Megatrend 2: *Innovation in the new economy will be driven by the blurring of disciplinary boundaries*. This program will provide you a unique mix of skills from international business, language technology, IT, marketing, management, e-business and intercultural communications.

- Megatrend 3: *Hyperconnectivity will redefine how people and organizations communicate*. This program will provide you skills to harness the power of hyperconnectivity to thrive in the networked economy.

- Megatrend 4: *Value systems will compete globally*. This program will equip you with the knowledge to "Think Global but Act Local."

- Megatrend 5: *Smart organizations will harness the power of "Collective Intelligence."* This program will teach you to leverage collective intelligence for global efficiency gains.

- Megatrend 6: *Economies of China, India and Brazil will be future engines of growth.* This program will teach you how to effectively tap these markets using the Internet.

The Executive Certificate in Web Globalization Management is a 110 hour self-paced online program for professionals involved with any aspect of conducting global business and e-commerce. Participants will learn strategies to leverage the Web for tapping global markets, conducting international e-commerce, managing virtual teams, developing network alliances, getting high search engine rankings, attracting and retaining global consumers, achieving culturally consistent global communications, and much more. Founded in 1910, the John Cook School of Business is the 15th oldest business school in the nation and has been a pioneer in recognizing the need for global business education. This vision has led to a highly recognized Executive Education Program (http://execed.slu.edu).

The magazine
MultiLingual
Language | Technology | Business

MultiLingual
d i g i t a l

Hello,

As publisher of *MultiLingual* and a co-sponsor of this *Guide*, I know you'll find it valuable in your work. To help you stay current with the ever-changing world of translation and localization, we would like to provide you with a complimentary subscription to the digital version of *MultiLingual*.

There's no obligation, and we won't send you a bill. Simply go to www.multilingual.com/promo and enter code BLSHD9. You'll receive an e-mail with a URL link when each issue becomes available.

Kind regards,

Donna D. Parrish

Donna Parrish
Publisher

www.multilingual.com

Notes

Contact Information - Contributors

	Lingo Systems - World Headquarters 15115 SW Sequoia Parkway Suite 200 Portland, OR 97224 Ph: 1-503-419-4856 • Fax: 1-503-419-4873 Toll Free: 800-878-8523 Email: info@lingosys.com www.lingosys.com	
	Lingo Systems - European Headquarters 25th Floor, 40 Bank Street Canary Wharf London E14 5NR UK Tel: +44 (0) 800-917-6564 FAX: +44 (0) 800-085-3911 Email: translations@languageline.co.uk www.lingosys.com	Chapter 22
	Language Line Services 1 Lower Ragsdale Drive, Building 2 Monterey, CA 93940 Toll Free: 1-800-752-6096 option 2 Email: info@languageline.com www.languageline.com	Chapter 21
	Author-it Software Corporation P.O. Box 200-273 Albany, Auckland New Zealand Ph: +64 (9) 915 5070 • Fax: +64 (9) 915 5071 Email: info@author-it.com www.author-it.com	Chapter 17
	MultiLingual Computing, Inc. 319 North First Avenue Sandpoint, ID 83864 Ph: 1-208-263-8178 • Fax: 1-208-263-6310 Email: info@multilingual.com www.multilingual.com	Chapter 2 and Foreword
	Common Sense Advisory 100 Merrimack Street Lowell, Massachusetts 01852 Ph: 1-978-275-0500 Fax: 1-978-275-0517 info@commonsenseadvisory.com www.commonsenseadvisory.com	Chapter 1

7th Edition

The Guide to Translation and Localization

Communicating with the Global Marketplace

"When someone who is not familiar with the translation and localization business asks me to recommend something to help them get a grasp of the industry, I immediately recommend this guide as the best resource. Language service buyers, investors, salespeople, and business owners can all benefit from the lessons and entertain themselves with the bloopers of *The Guide to Translation and Localization*."

Renato Beninatto
Chief Connector, Common Sense Advisory, Inc.

"With a solid mix of introductory information and expert insight, this book does an excellent job of 'translating the translation industry'."

John Yunker
President, Byte Level Research
Author, The Web Globalization Report Card

"As an academic in the area of Web Globalization, I found *The Guide to Translation and Localization* very useful for the students in the Localization Certification Program at CSU Chico when I taught there. The guide is an important resource for anybody wanting to familiarize themselves with current issues related to translation and localization. I hope to use the guide for participants in my Web Globalization/Localization classes."

Dr. Nitish Singh
Assistant Professor of International Business, and Program Leader for the Executive Certificate in Web Globalization at St. Louis University

Lingo Systems and MultiLingual Computing, Inc. are pleased to present the seventh edition of our award-winning book *The Guide to Translation and Localization*. Over the last ten years we have distributed more than 25,000 copies worldwide. Readers often tell us that they consider this book to be the definitive source for information about translation and localization. While accolades are always encouraging, our objective in creating the guide remains unchanged: to present clear, effective information to help localizers reduce cost, save time, and better understand global communication.

About Lingo Systems

Lingo Systems, a division of Language Line Services, provides customer-focused solutions in 170 languages for translation and localization, interpretation, quality assurance testing, multi-media and Flash engineering, and content management solutions. More companies trust Lingo Systems and Language Line for their language needs than any other provider. Shouldn't we be your partner, too?

For help with your localization needs, see **www.lingosys.com** or contact Lingo Systems at 800-878-8523 / **info@ lingosys.com**.

The Guide to Translation and Localization was published using the Author-it single-source content management system. See www.author-it.com for more information.

Author it ®