WRITING FOR TRANSLATION

GETTING STARTED: Guide

October/November 2009

PLANNING AND WRITING FOR TRANSLATION

OPTIMIZING THE SOURCE USING TRANSLATION MEMORY

ELEMENTS OF STYLE FOR MACHINE TRANSLATION

OPTIMIZED MT FOR HIGHER TRANSLATION QUALITY

CONTROLLED AUTHORING TO IMPROVE LOCALIZATION
Getting Started: Planning and Writing for Translation

Believe it or not, setting out to write lyrically beautiful copy for a manual or even the web may not be the most straightforward way to get to clear translation. These authors have some better ideas. Barb Sichel begins this Getting Started Guide with an overview on planning and writing for translation, and then Joseph Campo offers the findings from a project using a translation tool to find already-translated phrases to write the original copy. Ken Clark gives a short guide on writing for machine translation (MT), and Lori Thicke outlines why MT allows for quality translation in the first place. Ultan Ó Broin finishes things with a discussion on controlled authoring.

The Editors

Planning and Writing for Translation
page 3
Barb Sichel
Barb Sichel, director of business development at International Language Services, Inc., has over 25 years of sales, marketing and management experience.

Optimizing the Source Using Translation Memory
page 5
Joseph Campo
Joseph Campo, a senior technical writer at Dassault Systèmes SolidWorks Corporation in Concord, Massachusetts, has ten years of technical writing experience.

Elements of Style for Machine Translation
page 8
Ken Clark
Ken Clark, CEO of 1-800-Translate, worked previously as a journalist, screenwriter and speech writer for Japanese and American government officials.

Optimized MT for Higher Translation Quality
page 9
Lori Thicke
Lori Thicke is cofounder and general manager of Lexcelera (formerly Eurotexte), established in 1986, as well as cofounder of Translators Without Borders.

Controlled Authoring to Improve Localization
page 12
Ultan Ó Broin
Ultan Ó Broin, MultiLingual editorial board member and Blogos contributor, works for Oracle in Ireland. He has an MSc from Trinity College Dublin.
Documents and online communications are translated to achieve specific objectives. Your goal may be to execute a global communication plan, meet regulatory requirements, avoid liability or drive revenue by addressing target audiences in their native language. Whatever the outcome, you will need clear communication of a single message across all of the languages involved to get there.

Lately, cost considerations have become just as important as the accuracy of the translation. Consequently, writing for successful translation today involves planning your project so that you can convey your message within a reasonable budget.

Message and scope

First, and most obviously, decide what you need to communicate, and communicate it as simply and directly as possible. Determine what is most relevant to your target audience and what you must transliterate to achieve your particular objectives.

Take the time to think your project through from the perspective of the recipient, and do some research if you don’t know the recipient’s perspective. Translating everything you publish in English may not maximize the return on your translation investment. You might not have the luxury of translating every one of your product data sheets, for example, so focusing on product line summary brochures instead may be less costly. If it is beyond your budget to translate your entire 200-page employee manual, perhaps you can focus on only those critical policies most needed to protect your firm’s interests.

Some projects, such as catalog or website translations, may warrant the creation of abbreviated or revised versions for target audiences. Sections dealing with customer support or how to locate a sales representative, for instance, may need modification so that they are relevant in the geographic locale in which they will be used. Other types of projects require translation of ancillary materials that may not immediately come to mind. Technical documentation for large-scale industrial equipment may also involve translating warning labels and software user interfaces. Again, to save money, perhaps you can omit a section such as the corresponding parts list. If your customers can’t order parts in Japanese by calling your customer service line, why provide a Japanese parts list?

Understanding the intent and full scope of your project will enable you to plan your budget and work with your translator to determine the correct order in which to proceed. A phased implementation may be easiest to manage while allowing you to complete the highest priority requirements first.

Translation is a meticulous, skilled process similar in nature to technical writing.

Layout

For printed materials, properly planning your layout even before you start writing copy can greatly influence the ease and cost of managing your project. Quite literally, it pays to understand which factors affect the cost and quality of your translation. Then you can craft your presentation to achieve the desired outcome within your allotted budget.

A few things to consider are the choice of desktop publishing application and layout. If this is going to be a printed document with color plates, you might look at whether enough room is left for text expansion to accommodate any graphics. Text will expand in some languages and may contract in others. This has implications for the font sizes and page margins you select, as well as graphics. Chinese characters that need to be reduced to a 6-point font in order to fit on a page will be illegible. Also, check the graphics accessibility. Don’t plan to embed words into layer upon layer of graphics. Your translator may not be able to access them for translation at all or may be able to do so only at great expense to you. Plan to place your text labels beneath graphics rather than inside of them. Text must be “live,” that is, accessible independently of the graphics in order to be translated and reinserted in the same position.

The same concept applies to screen shots. Unless you translate your software first and provide new screen shots, the English copy locked within your graphics cannot be accessed for translation. If you must use preexisting graphics, your translator may be able to recommend solutions such as a reference table so that the reader can still understand your message.

Too often, project costs are unnecessarily high or the quality of the finished translation is compromised because translation was never considered when a document was originally created.

Your copy

Simple, straightforward text is easiest to translate. Say what you mean as concisely as possible. Word count is a key factor in the cost of your translation, so, if possible, keep sentences short and limited to a single idea. If English copy already exists for your pending translations, review and revise the content. Formal copy style with correct grammar, spelling and punctuation will be most easily understood by your translator. Consider also your audience’s education level and communication style and then select the appropriate tone. Instructions to a physician prescribing medication should be written differently than instructions to the patient taking the medication.

Avoid words with double meanings and references or metaphors that may not make sense in other cultures. Don’t rely on buzzwords, abbreviations, industry jargon, colloquial expression or humor. Create standardized text whenever possible. If you can reuse blocks of copy from one document to the next, you will save time and money on your translations and
ensure consistency across all of your written and online communications.

If your content is highly technical in nature or your industry-specific terms are prone to multiple meanings, supply your translator with reference material or glossaries for key terms. Links to websites or product catalogs can minimize the need for research during the translation process.

Some copy may not translate well or may translate into some languages but not others. Be particularly aware of this if you are creating ad copy or marketing materials. It is worthwhile to check with your translator early, before you have invested heavily in developing graphics or a tagline to accompany your corporate logo. Choosing the right words and the right images or colors for your presentation may make the difference between a seamless translation and one that falls completely flat with your target audience.

Acronyms should be avoided. The problem in trying to translate an acronym is that once you translate the theme word, the letters change and they no longer cross-reference to the supporting ideas you want to convey in your target languages. A native-speaking translator is a good resource for spotting things that won’t play well with your target audience. Basic localization — gearing your translated document to a particular country, region or target audience — is usually part of any well-executed translation project. Extensive localization, to the point of creative strategizing, however, is a specialized skill beyond the scope of typical translation projects. If you suspect your project requires an unusual amount of attention, check with your translator.

Provide only fully proofread, final copy for translation. Drafts are fine for budgetary reasons, but works-in-progress are unsuitable for translation and will leave yours prone to errors, inconsistencies and higher costs. If you intend to update documents later with new product models or next year’s catalog, the level of attention you devote to tracking changes and version control now will be well worth your effort.

Formatting
Locate your source files for older documents. This includes all of the desktop publishing and accompanying graphics files. Are they with your graphics design firm or archived somewhere within your organization? Your translator may not be able to replicate your formatting and graphics without them. Provide files to your translator in the same format you would like to receive back.

PDFs are fine for reference, but depending on the size of your document and the application used, having the source files available may significantly impact the time to quote your project, the cost of your project and the appearance of the final output. If you are working from hard copies or scanned documents, manual processes will have to be employed that will similarly affect your project.

Given the source files, most translation firms can replicate standard file formats, even for software code. How you present content for translation impacts cost, timeline and the ease of implementing your project. If you do any cutting and pasting at your end, have your translator provide a “post-format review.” This ensures proper text flow and the overall quality of your presentation before you print or post it on the internet. Costs for this service are usually nominal and can prevent potential embarrassment.

Formatting foreign character sets on your own can be a challenge, even for an experienced graphics person, and you may not have the right tool set. Languages such as Arabic that read right to left require special software versions and the ability to reorient everything on a page. It is best not to attempt this on your own.

If you are translating software for user interfaces, handheld LCD screens or similar uses, be prepared to answer questions about your ability to handle foreign character sets, space limitations and other factors that specifically affect these types of projects.

If you need to resize short translations to fit an ad or label, ask for an Adobe Illustrator EPS file that has been “outlined.” This provides the best of both worlds. It is locked down like a graphic to eliminate the possibility of introducing errors during formatting, but leaves flexibility for resizing. You can format the text to meet your needs, even for a character set that you may not have installed.

Lastly, use the right application for your project. Some applications play well with the automated tools employed by translation firms while others require a lot of manual manipulation.

Microsoft Word works fine for short documents, but FrameMaker may be a better choice for large manuals. If you use charts, live embedded links or manually inserted multiple carriage returns, the level of difficulty in working with your files for translation will increase, and this will impact your cost.

Timelines
Translation is a meticulous, skilled process similar in nature to technical writing. Though you provide the concept and the source files, your translator must take time to fully comprehend your meaning and find the best way to replicate the tone and content in his or her native tongue. Often there is research involved or requests for you to provide clarification.

Your project involves much more than merely translation. Numerous details are involved in preparing your files for translation, gaining commitment from the best qualified translators, proofreading, formatting and ensuring proper quality control. For multiple language translations, managing your project becomes even more complex. If you make a single change, it needs to be disseminated across teams of translators and proofreaders for each language.

Allow realistic timelines for your projects to be completed. A simple brochure may take several days, while a 300-page manual may take several weeks. Advise your project manager in advance if you must meet a specific deadline so that your project can be managed accordingly.

Partnering with a vendor
Since the quality of the translations you publish reflects on you and your organization, establishing a comfortable working relationship with your vendor is essential.

Carefully crafted branding strategies can be derailed in an instant by sloppy or inaccurate work. Even a single poorly chosen word can alter your intended meaning. And just imagine your customer purchasing a piece of equipment only to find that the documentation doesn’t make sense or that the table of contents doesn’t match the order of the text. You will rely on your translation vendor to provide you with accurate translations that are audience appropriate and delivered, print ready, within the specified timeframe. You should also educate yourself as to their quality processes and experience level with projects similar to yours so that you can move forward with full confidence.

While there is no single industry certification for translations, there are third parties such as TÜV of the American Translators Association that provide quality testing and auditing. It is perfectly acceptable to ask for credentials. In many cases, your own in-house quality policies or regulatory requirements demand that you do.

The Guide From MultiLingual
How many times have you written something and known that you wrote something similar, but can’t remember where it was or how it was written? If you could only find that text and replicate it, you would save money and time for your translation team by reusing already-translated text strings and would produce more consistent documentation.

This article describes a pilot project that tested a potential solution to this issue using translation memory (TM).

I hypothesized that if our technical writers could tie our authoring process into an English TM that contains already-translated text strings, we could find existing English text strings, reuse them on new topics and lower our translation costs. In effect, the documentation team would pretranslate their new English documentation to maximize matches against existing English text strings before sending topics to the translators who use the same TM.

We would use the English (source language) TM to improve the quality of fuzzy matches and reduce the number of words. Fuzzy matches indicate a percentage match of new or changed text against existing already-translated text. A higher percentage fuzzy match means the text string more closely matches existing translated text. The higher percentage the fuzzy match, the lower the cost to translate the text. Totally new text strings are the most expensive to translate, so I tried to reduce new words used. Because we translate into 12 languages, there is a great potential for cost savings.

After approval of the pilot project, I worked with my manager to schedule two months of project time. The translation team manager provided me with a TM tool — Trados, in my case — and I was ready to start the project after several days of training.

Project design

We use RoboHelp HTML to create online help and deliver multiple compiled help files (chms). I chose the main SolidWorks help to use in the pilot because it is our largest chm, with approximately 2,000 topics. I went back in time and created an English TM.

I collected 73 new and 39 changed topics that documentation had actually sent to the translation team during the SolidWorks 2007 development cycle. I used the Analyze tool in Workbench to obtain an original estimate of a full-cost translation. I also obtained an original estimate for a full-cost translation for the same topics from our outsourcing localization vendor, using German as the target language.

A dual monitor setup was essential to this project. On my right monitor, I opened Workbench and ran topics individually through the English TM to pretranslate them. On the left monitor, I opened the original HTML topic that had been sent to translation. When I ran a topic through Workbench, it provided a percentage match of the new text against the existing TM on a string-by-string basis. I used these suggestions to change the English source text in HTML on the left monitor and to improve the percentage of fuzzy match. I also paid strong attention to trying to reduce the number of new words.

After pretranslating each topic, I used the Analyze tool in Workbench to gauge and record the amount of savings for each topic. When I completed pretranslating all the topics, I calculated the costs and savings using the research data. I also obtained a translation cost estimate from our outsourcing localization vendor for the now pretranslated topics.

Results

In both Table 1 and Table 2, results show a modest reduction in per-language translation costs when comparing the original cost estimates to the cost estimates after using Trados to research the TM (post-Trados).

<table>
<thead>
<tr>
<th></th>
<th>Original cost</th>
<th>Post-Trados project cost</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>New topics</td>
<td>$4,807.64</td>
<td>$4,301.79</td>
<td>$505.85 (10.5%)</td>
</tr>
<tr>
<td>Changed topics</td>
<td>$1,957.97</td>
<td>$1,554.78</td>
<td>$403.19 (20.6%)</td>
</tr>
<tr>
<td>Grand total</td>
<td>$6,765.61</td>
<td>$5,856.57</td>
<td>$909.04 (13.4%)</td>
</tr>
</tbody>
</table>

Table 1: Cost estimate — vendor full-cost translation (includes translation, review and layout/DTP).

<table>
<thead>
<tr>
<th></th>
<th>Original cost</th>
<th>Post-Trados project cost</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>New topics</td>
<td>$4,463.57</td>
<td>$3,837.79</td>
<td>$625.78 (14.0%)</td>
</tr>
<tr>
<td>Changed topics</td>
<td>$1,322.25</td>
<td>$1,049.49</td>
<td>$272.76 (20.6%)</td>
</tr>
<tr>
<td>Grand total</td>
<td>$5,785.82</td>
<td>$4,887.28</td>
<td>$898.54 (15.5%)</td>
</tr>
</tbody>
</table>

Table 2: Cost estimate — Trados Workbench Analyze tool full-cost translation (includes translation, review and layout/DTP).
Details — new topics: I created charts to display the percentages of fuzzy matches in the original versus the post-Trados topics. The post-Trados topics (Figure 1) showed an increase in the number of 100% matches and a decrease in the number of No Matches. In terms of percentages, there was also an increase in the number of 50%-74% fuzzy matches. The remaining fuzzy match ranges were approximately equal to or less than the percentages of the original new topics.

- Total words reduced by 10% (2,028 words).
- 100% match increased by 439 words.
- No match reduced by 1,613 words.

Details — changed topics: Changed topics are existing topics with changes to already-translated text. These charts revealed a similar trend as with new topics. In terms of percentages, the post-Trados changed topics showed an increase in the number of 100% matches of about 10%. The remaining fuzzy match ranges were less than the percentages of the original changed topics. Overall, there is a greater percentage of 100% matches and a smaller percentage of no matches compared with the new topics.

Analysis

The cost estimates are within acceptable deviations that permit me to say that the Analyze tool results are defensible. I discussed the deviation with a senior employee in our research department. Standard deviations are complex to calculate and vary based on many parameters. When I provided the deviations, particularly for the new topics, the research employee felt that the 3.5% difference was within an acceptable deviation range (Table 3).

I then met with the translation manager to discuss the difference in costs between our outsourcing localization vendor and the Analyze tool. The translation manager confirmed that translation costs will vary depending on the vendor, the language, and the services provided. Having multiple variables makes it impossible to provide an exact cost estimate to fit all situations.

The translation manager informed me that only about 50% of our outsourced translation items require full-cost translation. She suggested we apply a different cost metric to the other 50% of our outsourced translation items; this metric is called raw translation, which includes translation of the text only. The savings in percent are similar to full-cost translation using the Analyze tool. Notably, for new topics, raw translation saved 14.1% while full-cost translation saved 14% (Table 4).

For the purposes of this pilot, I decided to split the difference between the outsourcing localization vendor savings of 10.5% and the Analyze tool full-cost translation savings of 14% and to use an estimated savings of 12.2% for new topics. This seemed like a reasonable compromise.

Savings projection

Outsourcing costs for a typical release vary from $100,000 minimum to $400,000 maximum, depending on how many new products and services requiring localization are added to our suite of products, their length, and the number of languages supported. If the process was applied to all new documentation that we send to translation for outsourced localization, an estimated cost savings of 12.2% (between $12,200 and $48,800) would be achieved (Table 5).

<table>
<thead>
<tr>
<th></th>
<th>Original cost</th>
<th>Post-Trados project cost</th>
<th>Savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>New topics</td>
<td>$2,409.98</td>
<td>$2,071.03</td>
<td>$338.95 (14.1%)</td>
</tr>
<tr>
<td>Changed topics</td>
<td>$ 815.93</td>
<td>$ 629.67</td>
<td>$186.26 (22.8%)</td>
</tr>
<tr>
<td>Grand total</td>
<td>$3,225.91</td>
<td>$2,700.70</td>
<td>$525.21 (16.3%)</td>
</tr>
</tbody>
</table>

Table 4: Cost estimate — Trados Workbench Analyze tool raw translation (includes translation only).

<table>
<thead>
<tr>
<th></th>
<th>New topics savings</th>
<th>New topics deviation</th>
<th>Changed topics savings</th>
<th>Changed topics deviation</th>
<th>Grand total savings</th>
<th>Grand total deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vendor</td>
<td>10.5%</td>
<td>33.3%</td>
<td>20.6%</td>
<td>0%</td>
<td>13.4%</td>
<td>17.9%</td>
</tr>
<tr>
<td>Analyze tool</td>
<td>14.0%</td>
<td></td>
<td>20.6%</td>
<td></td>
<td>15.8%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Full-cost savings comparison/deviation between vendor and Analyze tool.
Changed documentation is typically localized by our in-house translation team, so for us there will be no “cost savings” per se. However, the translation team would experience a time savings of between 20.6% and 22.8% because of the increased quality and number of matches as well as the reduced word count.

Pilot project conclusions

Using a TM tool is viable in pretranslation only if we consider its value in increasing the consistency and quality of documentation. I could not justify using the tool on just a cost-savings basis alone.

Savings were achieved by both reuse of existing text and aggressive word-count reduction. However, the anticipated translation savings only partially offset the cost of the skilled writer’s time in editing. I spent approximately 30 minutes per topic using my TM tool.

For the 73 new topics, I spent approximately 37 writer hours. Using $50 cost per writer hour, I spent $1,850 in time to achieve only $625 savings in outsourcing costs. Labor costs were triple the savings achieved, for one language. Actual cost savings are only achieved when factoring in that we translate into 12 languages.

Savings = total cost savings ($7,500) – time spent ($1,850) = $5,650. If the TM tool were used to only search for reusable text (no word reduction), the results would be even less impressive (estimated 2.4% savings in outsourcing costs).

Beyond the case study: related research

I queried translation experts as to whether any similar projects had been undertaken. Authoring memory tools have been around for over ten years. An article by Jeff Allen in 1999 discussed how authoring memory could be used in conjunction with controlled language to aid in translation (www.transref.org/default.asp?dscsrc=/u-articles/allen2.asp). The new Author-it product, for example, uses fuzzy logic matching within a content management system. I contacted Nabil Freij, president of GlobalVision, and according to him, this pilot project was a unique approach. The key to reducing localization costs is to reduce word count. Some companies are implementing controlled English to reduce word count, increase the 100% matches, and also to transition to machine translation (MT). According to Freij, “MT engines can perform better under restricted and controlled vocabulary.” In his experience, “most tech pub writers do not want to deal with localization issues during the authoring stage, they are simply too overworked. . . . We often can’t get them to edit their work, let alone reduce the word count or make it consistent.”

I have been following the progress of the SDLX AuthorAssistant (SDLXAA) product, which seems similar to my pilot project. SDLXAA lets writers write, then runs the document against a TM to offer suggestions for improved matches and reuse. According to Sue Blaisdell, information architect at Avaya, “with AuthorAssistant, you can connect to TMs for your project, and it will display 100% and fuzzy matches to the writer. It also gives the writers insight into the way that changes they make in their English content affect the localization costs.”

This pilot project indicates that translation cost savings can be achieved using TM, but at a cost in labor and time. With usage, writers would become more proficient with the system and save time. Your company would have to be ready to absorb license and time costs. If you are going through a major restructuring of your documentation, perhaps upgrading to XML, this might be the perfect time to examine your documentation in detail with your translation costs in mind.

One benefit I found was that while using my TM tool, I was fully focused on reducing word count because I kept translation as my main focus. Word reduction is hard to achieve in normal writing mode because the technical writer is normally not thinking about it. According to Freij, “verbosity is the enemy. It pays to be concise and straight to the point, eliminating unnecessary text when localization is imminent. When writing technical documents, remember that simplicity is also very much desired by the end-user.” It is also important that your TM be as clean as possible.

What writers need is a TM tool that runs side-by-side with an authoring application and can semi-automatically offer suggestions on how to better match new text to the existing TM. The development of SDLXAA and Author-it’s new application should give us hope that tools are becoming available to bring technical writers and translators closer together to achieve cost savings by leveraging valuable memory resources.

--

<table>
<thead>
<tr>
<th></th>
<th>Outsourced localization cost savings</th>
<th>SolidWorks translation team time savings</th>
</tr>
</thead>
<tbody>
<tr>
<td>New</td>
<td>$12,200 to $48,800</td>
<td>n/a</td>
</tr>
<tr>
<td>Changed</td>
<td>n/a</td>
<td>20.6% to 22.8%</td>
</tr>
</tbody>
</table>

Table 5: Annual estimated savings if Trados is implemented for all new and changed documentation.

UPCOMING EVENTS

LOCALIZATION WORLD
CONFERENCE & EXHIBITS

2009 Know-how for Global Success

- October 20-22
- Hyatt Regency Santa Clara, Silicon Valley, California

2010

- 7-9 June
- Hotel Maritim proArte, Berlin, Germany
- October 12-14
- Bell Harbor Conference Center, Seattle, Washington

info@localizationworld.com

www.localizationworld.com

October/November 2009 • www.multilingual.com/gsg
Elements of Style
For Machine Translation

Ken Clark

We have entered the Machine Translation Age. Demand for human translation is still increasing dramatically—or was until this year—but the vast majority of the world’s translation is now done by computer. And the vast majority of machine translation (MT) transactions is completed using free online translation tools such as Babel Fish or Google.

The result usually leaves much to be desired, and there’s not much you can do about it when you are translating someone else’s content, particularly if you don’t understand the source language. But you can dramatically improve translation of content you write yourself and share with others in a foreign language, without using the special software and workflows of powerhouse automated translation systems. Just a few simple writing tricks can make a dramatic difference in MT quality. It’s not controlled language, but language control.

Writing clearly, whether for man or machine, is always a struggle (at least for me), and the dim machine minds of the translation tools are unforgiving when it comes to bad composition in source. Unlike us humans, MT tools have no sense of context, no appreciation of an author’s intent and definitely no sense of humor.

With Strunk and White’s The Elements of Style as inspiration, here’s an abbreviated guide to good English style for improved MT.

- Use short sentences. Keep it simple. Cut the clauses. Ditch the sentence fragments. Simple sentences and grammatical structure (subject-object-verb) are the only way to go.
- Avoid ambiguity, as in “I saw her duck.” Well, which is it? A duck that quacks that belongs to her? Or was she avoiding a flying object? Look for multiple meanings when proofing. Good luck. If you don’t find it, your MT tool may just do it for you.
- Remove extra words. Editing out unessential phrases and extra words will make for a simpler, better translation. Since the algorithms have fewer translation variables to wrestle with and better style with fewer words in the translation, it will also be more accurate.
- Don’t remove necessary words, and don’t go too far with editing. In English we drop a lot of words when we write, especially when writing informally. Keep those articles, prepositions, pronouns and so on where the machine can find them. English speakers are able to fill in the blanks and fully understand—not so when the reader is a translation engine.
- Misspelling does not compute. A misspelled word will not translate—end of story, end of translation.
- Ditto on punctuation. One accidental period can completely change the meaning of a sentence and trash your translation. Spell-checking and proofreading after you write and before you translate are pretty basic quality assurance steps.

The simplicity and clarity of expression demanded by MT tools would meet the approval of Strunk and White, I hope.

- Slang is so like, whatever. No slang and no jokes for MT. Irony is the first thing lost in MT. Stay earnest and formal. That’s why pithy headlines and snappy newspaper copy so often translate badly with these tools. Rule of thumb: Good MT style comes in one flavor . . . plain vanilla.
- Use “Do not translate” coding. Some MT tools will allow you to place code around a word or phrase, which allows the word to pass through the engine without getting translated.

- Check your translation. So after closely studying and applying all these rules before translation, how can you know if your MT output makes any sense? Translate the output back using an MT tool. That reverse translation may help you to spot the most glaring errors. Recast those problem sentences in English and see if the back translation gets any clearer. Don’t expect miracles here. But it may be some comfort to know that the original translation is better than the back translation.
- Keep source and target together. No garbage in the MT tool, less garbage out. But garbage there will be. That’s why we like to keep a copy of the source with the target translation to create a bilingual output so that those errors can be spotted and corrected later if need be.
- Identify MT. Avoid blame by giving credit. Letting people know that you used a machine to communicate with them allows them to read with caution, and keeps them from feeling you’ve been short-changed on a real translation.

On the writers’ craft

Using a little bit of discipline to prepare content for MT extends the functionality of these tools for people busily engaged with others in multiple languages.

Writing for MT, just like writing for human translation, is good writing practice. Translation has a way of highlighting communication errors that are invisible or ignored in just a single language. The simplicity and clarity of expression demanded by MT tools would meet the approval of Strunk and White, I hope.


I’ve still got a dog-eared, ratty, old copy on my desk, where it shall remain.
Everyone knows machine translation (MT) has enormous potential for dramatically reducing translation cost and increasing speed. But who thinks of MT as a way to improve quality?

ISO 9001-certified for the last decade, my company’s quest for quality has unexpectedly led us to MT. Along the way we’ve developed and tested a number of different processes for MT and discovered that correctly optimized MT can actually improve quality — and for less cost and with higher rates of productivity. Under the right conditions, MT actually breaks those compromises we’ve come to accept in the traditional localization paradigm. You may want price, speed and quality, but here’s the kicker: you only get to pick two out of three.

MT can offer all three. However, the truth is that for most people, quality MT is still an oxymoron. And who could blame them?

MT: always five years from perfection

Just about any of us with an internet connection has had first-hand experience with MT. We have probably used SYSTRAN to translate an e-mail or ProMT to give us the gist of a web page. We may have conversed with someone in another language via Google’s translation center, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online or, more recently, read Wikipedia in Thai thanks to Asia Online.

Along the way, MT will have amused us with its inadvertent twisting of human language.

Most people would agree that “out-of-the-box” MT is far from what it is supposed to be: fully automatic quality translation (FAQT). This has been the promise held out to our industry since the very first MT system translated 49 Russian sentences into English using a 250-word vocabulary and six grammar rules. Fifty years later we’re still waiting. As Hans Fenstermacher of Translations.com says, “MT has been five years from perfection since 1952.”

It could be that our overwrought expectations for MT partially explain the slow uptake of MT by the translation industry. Against the benchmark of FAQT, MT is sure to disappoint. For those resigned to the lack of quality with unoptimized MT, there’s always the unfortunately named FAUT (fully automatic useful translation). FAUT is essentially “gisting” translation, which is a more or less accurate approximation of the source text.

Today, gisting is overwhelmingly the use to which MT is being applied and accounts for even more words translated than by humans. If the claim that MT translates more than humans seems outrageous, consider that an estimated 30 million e-mails are translated by MT every day.

For internauts, instantaneous gisting (gist-in-time) provides a basic understanding of an e-mail or a website. In the corporate space, gisting is used for legal discovery, for patent or technology searches, or to identify parts of larger corpora that merit being translated by a human. But how much gisting do we humans really need? Not much, as it turns out: for all the profusion of free, software-as-a-service and off-the-shelf MT solutions, commercial translations, which need more than gisting quality, are by and large assured by humans. For the vast majority of corporate needs, MT is staying on the shelf.

If FAQT is still “five years away” and FAUT is simply not that useful after all, the question is whether to wait for MT to catch up to our aspirations for it or to invest in processes that can optimize the MT we have today.

How MT improves quality

Once we stop waiting for quality MT to emerge fully clothed from the loins of a research and development lab somewhere, we can start to see MT for what it is: an efficient solution that can assist human translators by taking out a large part of the drudgery in translation.

The reality we are seeing every day is that for technical translations ranging from software to manuals to catalogs, quality MT is achievable. But like any relationship, you have to work at it. In fact, correctly optimized MT — that’s the “working at it” part — paired with human post-editors can actually improve quality. How could this be possible?

In the first place, correctly customized MT (customizing MT engines is a skill in itself) removes terminological inconsistencies. If the source document always uses the same term, so will the MT engine. This resolves the real problem of teams of translators working on the same project but employing divergent terminology. Across a large project, MT can also ensure a more consistent tone, with less stylistic discrepancies. Furthermore, MT removes that human element of non-quality: omissions. Enforced, validated terminology, consistency and completeness are MT’s strengths. But what about mistranslations? There’s no question that MT delivers more of its fair share of sentences that mangle the meaning of the source text.

This is where the post-editors come in. Working on a bitext format, a post-editor correcting MT output will frequently scrutinize texts more carefully than a reviewer working on human output. On large-volume localization projects, T + E + P (translate + edit + proof) as a process may be interpreted differently by different language service providers. T + E + P on a million-word project may consist of T + a sampling review of 20-20. The source text may or may not be consulted at the same time.
MT affords you no such luxury. Because MT can and does go completely off the rails from time to time, each and every segment must be examined in bitext format and approved or rewritten by a human posteditor. If only every translation received that type of attention!

This process for review and correction, if properly managed, should not only catch and fix the errors, but should also yield an accounting of what changes need to be made to the MT engine itself. This goes to the heart of any good quality system, such as ISO 9001: ensuring quality at the source — that is, catching errors at the beginning rather than correcting them downstream — and, crucially, instituting processes for continuous improvement.

Correcting systematic errors and then feeding these corrections back into the MT engine is what we call “the Virtuous Circle of MT Quality.” This, too, is an integral part of the optimization process.

What quality do you need?

But what quality is good enough? Any good process defines its quality expectations up front, and working with MT is no exception.

MT quality has been measured by the wrong yardsticks to the detriment of the elegant solution that MT can be when matched to the type of result needed. The question is not whether MT is “better” than a human translation on a given text. Rather, the question is what quality is necessary for a particular project and what process — human only, human + translation memory (TM), human + TM + MT — will best allow you to achieve that exact level of quality.

The 2008 version of the ISO 9001 standard introduces the idea of customer-defined quality to the international norm. This is an important distinction to make. Accuracy, consistency of style, correct terminology, spelling and punctuation, and completeness are all inarguably elements of a quality translation. But how much quality is required for a given situation? “Doesn’t read like a translation,” for example, is the type of quality that a marketing translation would need to have in buckets. We may not have a specific metric for defining marketing quality, but we sure know when it’s not there! But what about software? A catalog? E-learning courseware? A knowledge base? This is where the quality question begins to get more nuanced.

For software, quality may be defined as accurate, understandable and rapid enough for simphs. For a catalog, correct terminology on each of thousands of items is primordial. For courseware, the material needs to promote learning. For a knowledge base, customers need to be able to resolve their problems without further recourse to the help desk staff.

Since MT allows you to calibrate the human effort (linguistic training, post-editing) that you put into achieving the quality levels you need, setting quality requirements in advance is an essential step. The example of online help and knowledge bases above demonstrates the importance of customer-defined quality. It’s well known that human reviewers will often designate only extremely high quality as acceptable. However, when the choice is between an imperfect translation and no translation (information available only in the original language), customers themselves weigh in heavily in favor of raw — that is, fully automatic — MT.

Don DePalma of the Common Sense Advisory says, “Whether it’s FAQT, FAUT, or perfectly rendered output, the biggest decision that companies will have to make about machine translation is whether any of those are a worse alternative than no translation at all. Given the enormous volumes of content that companies and government should make available for other markets, for me and many of the organizations that we talk to, the quality question is ultimately a non-issue. What we call the ‘zero translation’ option of doing nothing means no information, service or customer satisfaction at all.”

Customers also report that support articles translated by MT are just about as effective in solving their problems as human localized content and at a price far below what human translations would cost.

This is not about depriving translators of work. Human translations would not have been economically feasible for the hundreds of thousands of knowledge base articles in various languages — including Chinese, Japanese, Portuguese, French, German and Spanish — that Microsoft publishes online. This would have required an initial outlay for approximately $30 million per language, according to Microsoft itself, not including weekly updates. Instead, Microsoft chose its own hybrid MT system to translate content that would otherwise not have been translated. Measuring the results, the company found that across all languages, MT helped solve customer problems on average 23% of the time. This figure may seem low, but it’s only slightly below the success rate of 29% for human translation.

Microsoft concluded at a presentation to the 11th Machine Translation Summit in Copenhagen, Denmark, in September 2007 that “customer satisfaction numbers for machine translated articles is comparable to and sometimes exceeds original English!”

Optimizing MT

Regardless of the quality level MT is to achieve — publishable quality or simply understandable quality — unoptimized MT is just not up to the job. While some sentences coming out of untrained MT engines may be stunningly good, others will be pure gibberish. And without effective training, there is no way to ensure that the terminology you want will be consistently applied by the MT engine.

Training, then, is the secret sauce of good MT, even more important than what system you choose, whether rule-based or statistical (see sidebar). This is also one of the areas that requires the greatest investment. For statistical machine translation (SMT) systems, this training involves not only extensive corpora of bitext (think in terms of millions of segments), but also glossaries and monolingual texts. The more the better. Imagine Steven Spielberg’s little alien, ET, saying “Need more data.” That’s SMT in a nutshell.
Rule-based versus statistical MT

There are two major streams in MT technology: rule-based MT (RBMT) and statistical MT (SMT). These two methods, espoused by various MT technology vendors, represent two different routes to the same place.

The earliest systems were rule-based, among them SYSTRAN. For the development of RBMT systems (SYSTRAN, ProMT, Lucy), various languages were broken down into their parts of speech and grammatical rules were hard coded, along with dictionaries. An RBMT system would never say un noir chat but un chat noir, coded, as it is, with the knowledge that adjectives follow nouns in French. Exceptions such as une vieille dame would also be coded in the system.

SMT, on the other hand (Google, Asia Online), uses an algorithm to parse vast numbers of bilingual sentences (preferably in the millions) in order to extrapolate relationships, including word order. Un chat noir would appear as the translation of a black cat if it had seen that in the training phase. However, blissfully ignorant of the rules of grammar (with the exception of Asia Online), SMT would be likely to incorrectly translate a green cat as un vert chat because it wouldn’t have encountered any green cats — unless trained on Dr. Seuss.

Both RBMT and SMT systems have their advantages and disadvantages. Both are capable of delivering accurate, fluid sentences, depending on how they were trained. Both can also deliver utter gibberish — again, depending on how they were trained. RBMT wins the day when you don’t have millions of words of training corpora; SMT is the victor when it comes to adding a new language pair, a major multiyear undertaking when preparing an RBMT system. Hybrid systems such as SYSTRAN’s are capable of bridging the gap between RBMT and SMT.

Testing will provide information on the level of fuzzy match that should be discarded in favor of MT segments. However, it’s usually useful to make sure that new MT segments are identified as such to distinguish them from validated TM segments.

Long, convoluted sentences do not lend themselves to MT, no matter how well trained the system is.

The capacity of MT to function as a standalone will depend on the quality required and on how well the engine is optimized through stringent training, ongoing maintenance, controlled authoring and so on. But for publishable quality, human post-editors are essential.

In this regard, MT can be seen as just another tool in the translator’s toolkit, much like any CAT tool, albeit one that’s more complex and expensive to set up. In optimizing MT, post-editors need to be trained in post-editing techniques, and they need to know what level of quality is expected. Besides post-editing, other post-production optimization techniques include use of QA tools, automatic post-editing through regular expressions, text normalization, updating of the TMs and so on. And above all, it is essential that there be ongoing tuning of the engine with new and modified terminology and error corrections in a continuous, virtuous cycle of feedback and improvement.

If all these processes, from pre-production to post-production, are instituted to optimize MT output, what kind of quality can be expected? Recently one of our clients, a major software publisher, noted in the report “Leveraging a crisis for innovation (or never let a good crisis go to waste)” that “contrary to all expectations, using MT in [our company] has improved the translation quality . . . with the reviewer commenting ‘It was nearly 9 — it was the best translation of coursework I ever read.’”

It has long been believed that buyers of translation services must compromise. In the traditional localization paradigm, if you want speed and quality, you have to compromise on price; if you want speed and price, you have to compromise on quality. MT is often associated with a compromise of quality in favor of cost and turnaround improvements. However, the reality is that correctly optimized MT can break these compromises by offering faster throughput, lower costs and higher quality. But you have to work at it.
CONTROLLED AUTHORING TO IMPROVE LOCALIZATION

ULTAN Ó BRION

Controlled authoring, broadly speaking, is the process of applying a set of predefined style, grammar, punctuation rules and approved terminology to content (documentation or software) during its development. Many companies offer some form of guidance to their content developers, either through tools or more ad hoc means, of course, so this may not seem at all remarkable. In the last few years, however, innovations in linguistic processing technology and its commoditization indicate that controlled authoring holds great potential for anyone seeking a tool-driven approach to maximizing returns from the localization process. This has particularly paved the way for the adoption of cost-effective machine translation (MT).

Controlled authoring and languages are complex, so this article concentrates on the localization-related aspects of introducing controlled authoring into an organization that must localize its content.

Controlled authoring itself is frequently conflated with other parts of the overall content development process, notably that of content management, a separate but contributing function. Isolating the non-technical essence of controlled authoring is made all the more difficult by the range and interplay of tool functionality offered by various vendors. Whereas seemingly subtle distinctions do not always make a great deal of sense from an overall business process engineering viewpoint, it’s important to understand from a localization perspective just how controlled authoring technology works. For example, if the storage of objects in the content management system (CMS) allows reuse at a level higher than the translation memory (TM) segmentation does, localization saves are limited. It may be more helpful from a business requirements position to regard controlled authoring as an information quality process that consists of many different parts: data mining for rule and terminology research and creation, new terminology harvesting and rule development, reuse management, reporting on quality, and so on rather than purely approved rule and term application during the actual text-editing phase.

Controlled languages

It’s not uncommon for organizations to have no serious control over their content style rules and terminology or to rely on manual processes, combining in-house guidelines with the commonly applicable rules and recommendations of sources such as The Chicago Manual of Style, while working off spreadsheets of terms and applying simple checks for consistency and using human editing to meet their “quality” requirements. For some this is acceptable; however, it is hardly a scalable, enforceable or measurable process. We’ve all seen the waste of many possible opportunities for localization efficiencies — let’s save the content development efficiencies for another audience — because manual enforcement and voluntary uptake of authoring guidelines allow for a good deal of subjectivity in interpretation and application. Controlled authoring is much more objective as the selection, application and enforcement of such guidance is programmatic. The application of rules “controls” the authoring, so to speak, allowing content developers to avail of the rules directly through the authoring user interface: looking up alternative words, phrases and terms, reusing already written phrases, harvesting and storing new ones, and reporting on the content’s compliance with the rules immediately or afterwards.

The origins of the controlled language concept are far from the needs of modern day localization, rather being designed to improve comprehension of the source language by simplifying matters for nonnative English speakers of English (“human orientation”) or computers (“machine orientation”). Often, these nonnative readers worked in the maintenance and service field, and were targeted by probably the best-known iteration of a controlled language: ASD-STEP200 Simplified (Technical) English. The genesis of the controlled language concept can be traced back to Ogden’s Basic English from the 1930s and established over the years through such developments as Caterpillar Technical English, Nortel Standard English, the Plain English Campaign, GM’s Controlled Automotive Service Language, Global English and so on.

The introduction of structured authoring through SGML and later XML, along with more innovations in linguistic processing and database storage, allowed for the development of and application of targeted rules to meet customer requirements driven by content type and market, reflected by the ability to now apply a controlled authoring process through common authoring tools such as Microsoft Word, PTC Arbortext Editor and Adobe FrameMaker through plug-ins.

The use of an approved set of terminology, where each term has only one meaning in that context — consider the different translations for the out-of-context word job, for example — and clear and enforceable authoring rules allow writers to achieve a high degree of consistency in the source texts they create, not only in the words and terms they use, but how they use them. Consistency in constructing phrases, along with eliminating complexity, ambiguity and verbosity, is the key to maximizing TM use and MT potential (and large efficiencies on the production side).

What might these controlled language rules entail? Well, the number can vary, could be as many as 10 to between 50 and 100, but typically might relate to standardized spelling, length of sentence, number of clauses, use of active versus passive, simplifying tenses, rules for noun phrases, modifiers, syntactic cues, past participles, gerunds, avoidance of Latin phrases, slang and so on. I recommend Jon R. Kohl’s The Global English Style Guide if you need a valuable starting point and reference material for possible rules as well as Sharon O’Brien’s “Controlling Controlled English” paper (www.mt-archive.info/CLT-2003-Obrien.pdf) for recommendations on the rules central to content intended for MT.

Naturally, the rules vary by content type and audience. Gerunds may be acceptable in headings, but not main text without qualification, delimiters may not be required...
<table>
<thead>
<tr>
<th>Feature</th>
<th>Solution 1</th>
<th>Solution 2</th>
<th>Solution 3</th>
<th>Weighting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Price</td>
<td>NLP-level verification of terms, grammar and style according to our requirements</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prompting of writer to reuse of existing segments from CMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scalability (multiple users, concurrent users, performance)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Easy maintenance of rules by existing, in-house resources</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of existing rules from style guide that can be automated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Integration with existing translation glossaries and exchange formats</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New terminology harvesting</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic rule set supports translation memory requirements</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic rule set supports machine translation readiness</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic reporting on quality in batch and single mode</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interactive quality assurance through editing environment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allows prioritization of rules for grandfathering of content</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supports multiple rules and terms by content type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plug-ins and integrations for existing authoring tools</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Automatic indexing capability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Customer references include MT and TM savings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open standards or proprietary architecture</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Established user group and conference</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Global 24 x 7 support</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Business requirements can be weighed against a variety of solutions.

on software strings but are required on messages or documentation, and so on. Thus, solutions that allow for forms of semantic checking have an advantage.

**Controlled authoring solution: which one?**

Commercially available controlled authoring technology solutions range from the more sophisticated, scalable technology-based solutions based on advanced linguistic processing to less complex content management-based offerings, “methodologies” and combinations of same. Possible options include acrolinx IQ, IAI CLAT, Author-it, Smart MAXit, Boeing Simplified English Checker, SDL AuthorAssistant, Shufra and more. For those interested in researching a controlled authoring option, possible sources of information are IDC reports, ELDA, *International Journal of Language and Documentation*, CLAW proceedings, Localisation Research Centre publications, DCU papers from Sharon O’Brien and MA research by Patrick Cadwell (www.localization.ie/resources/Awards/Theses/PatrickCadwell_Thesis.pdf), the publications of Jeff Allen (www.geocities.com/controlled_language), the MT, Localization Professional, and Information Quality groups on LinkedIn, and so on.

The decision as to which controlled authoring solution to adopt is driven by business requirements. Localization teams should ensure they’re involved in the identification of these, so come armed with facts and figures for the business case. Initial business requirements when applied to a range of solution possibilities might look something like Figure 1.

Requirements vary by organization, naturally. Prioritize and weight each point before making a decision among competing alternatives.

**Benefits to localization**

The clear benefits of controlled authoring in the localization space are derived from the improved source quality making content easier to translate by humans and machine. It’s a fundamental recognition that the basic internationalization concept of assuring translatability and high-quality source content results in greater savings accruing to the organization at the localization stage than trying to continually negotiate lower prices with vendors or praying for a quantum leap in translation technology to turn garbage source into localized gospel. Efficiencies are magnified in a one-to-many relationship as the number of languages translated increases.

Controlled authored content is consistently expressed in an understandable way. This results in translators not needing clarifications, maximizing TM matches, eliminating the need for terminology creation after localization starts, and providing texts more easily processed by MT, cutting down on post-editing needs and recalibrations. Volumes too, are generally smaller, reducing cost and time-to-localize per se.

Bear in mind, however, that these savings are a function of the rules created, as well as how and when the texts are translated. Overarching internationalization rules also impact the efficiencies as well as the technical review of the source text by domain experts. If a switch should be documented as being “off” instead of “on,” then don’t expect controlled authoring to eliminate any language version testing issues.

Do you need controlled authoring technology in order to use MT? The simple answer is “no.” But if you need a scalable approach to ensure your source text meets realistic MT business requirements by providing easily processed source text that minimizes the need for post-editing, thus making MT cost-effective, then controlled authoring technology is a must-have. Moving past the “writing for translation guidelines” approach is the way to go here.

**The business case for controlled authoring: the big picture**

It’s often said that the biggest risks to the introduction of controlled authoring, other than the cost (nontrivial even at the best of times), is the political. The term controlled authoring itself must be found guilty on all counts of contributing to the problem of user acceptance as it conjures up images of mass layoffs, stilted, boring texts, loss of control by authors, inducing an immediate negative reaction, mostly based on understandable fear and ignorance, frequently exacerbated by a belief...
that controlled authoring can somehow offer automatic creation of content, and a narrow focus on just localization benefits. There are ways of dealing with these issues too, beyond the scope of this article.

In general then, beyond the clear hard-sell on the TM and MT front, localization cost and time-to-market savings, localization teams can emphasize the quality aspects of the source content for native speakers too — superior user experience, consistent terminology, less support calls, improved accessibility and so on. Leverage the global user experience, not just the localized one. It should be pointed out there are controlled authoring solutions for Japanese, German and so on, so do not assume it is an English-only concept now, whatever the origins.

Introduction and changing processes: localization's role

Introduction of controlled authoring requires a serious management decision as to timing, not least the provision of a significant budget. However, research would indicate that using pilot projects to develop the process as well as achieve maximum buy-in by the stakeholders in the process is key, as well as using training techniques that rely less on computer science and linguistics but more on content development approaches. In general, localization groups might consider the following with faced with the opportunity to introduce controlled authoring:

- Identify a localization strategy for TM and MT tools and how controlled authoring business requirements fit into this.
- Help kick-start the controlled authoring process of adoption and pilot projects by providing rules and terminology already harvested to the implementers of the technology.
- As the creation of rules and terminology are central to controlled authoring and to the impact on localization tools, then it is critical that localization groups remain visible and active as stakeholders in their development and maintenance over time.
- Recognize the best kind of texts — large volumes of structured, technical, procedural texts such as software and user assistance strings or online documentation. These texts require a consistent user experience between components. Seeking a controlled authoring solution for a few thousand words of marketing material would not be a strong business case!
  - Prioritize rules. Decide which ones are more important to you than others. Aim for automatable and therefore measurable ones. For example, a rule called "one strong idea per sentence" is not automatable, whereas repeating the noun instead of switching it for a pronoun or checking for the passive voice is.
  - Look for leverage points between localization and authoring teams. Many rules for localization maximization are obviously ones that should be applied to text even if never intended for localization in the first place. Other, more "severe" MT rules may not be optimal for the source language depending on the user experience required. For example, text intended for mobile applications may be fragmented, clipped, dropping articles and so on for user experience reasons. Be prepared for compromise. Err on the side of user experience trumping localization unless it’s a complete showstopper.
  - One particular challenge to the introduction of controlled authoring can come from localization groups themselves — the disruption of TM match rates for previously localized content. This requires careful management. Solutions include the introduction of controlled authoring on new content yet to be localized, phased introductions based on content that is going to change anyway, grandfathering of content that has shown little change over years, or a reassessment as to how a one-time hit on localization assets results in longer term cost savings, time-to-market improvements and quality uptake.
  - Localization group input to the rule creation process must be matched by an evaluation of the localized source output, too, iteratively maximizing returns through the fine-tuning of rules. An MT pilot makes a fine adjunct to a controlled authoring pilot. Provide content development teams with the feedback, qualitative and quantitative.

Acknowledgements

Publicly available sources from the following were used in this article: Patrick Cadwell (DCU), Sharon O’Brien (DCU), Jeff Allen (Translations.com), Uwe Muegge (Muegge.cc), Jon Kohl (SAS), Andres Heuberger (ForeignExchange Translations) and Fred Hollowood (Symantec).
Creating a Dialogue with the World

Our network of 500+ skilled professionals, working in over 50 world languages and numerous areas of expertise, provides you with the precision and accuracy needed in today’s global marketplace.

- Interpretation and translation services
- Competitive pricing
- 50+ languages
- Free, no-obligation estimates

Tennessee Foreign Language Institute
Nashville, Tennessee USA
info@tfli.org • www.tfli.org

Translation Services into 70 Languages

We provide translation services into 70 languages using the most modern technology for clients throughout the whole world.

- CEET provides translation, proofreading, localization, DTP, interpreting, voice-over and cultural consulting in all major world languages with special expertise in Central and Eastern European languages.
- We approach all projects with respect to customers’ needs and the cultural uniqueness of each country because we believe the language of your firm communicates who you are to the audience.

SDL TRADOS Technologies

SDL TRADOS Technologies is a division of SDL, the world’s largest provider of technology solutions for global information management (GIM), which benefit corporations and institutions, language service providers and freelance translators worldwide.

SDL has over 170,000 software licenses deployed across the translation supply chain and has demonstrated proven ROI in over 500 enterprise solution installations. SDL delivers innovative software products that accelerate global content delivery and maximize language translation productivity.

SDL
Berkshire, UK
www.lspzone.com

CEET Ltd.
Prague, Czech Republic
info@ceet.eu • www.ceet.eu

SDL
Berkshire, UK
www.lspzone.com

Clear Words Translations
Córdoba, Argentina
info@clearwordstranslations.com
www.clearwordstranslations.com

Clarity and Efficiency

With a vast network of professionals worldwide, we provide reliable, customized language solutions in Spanish and Brazilian Portuguese.

Our services include localization, translation, interpreting, desktop publishing, and project management solutions that enable our clients to increase revenue and create effective communication channels with their audiences.

Our adherence to on-time deliveries, fair pricing and fast turnaround makes us a language service provider our clients can trust.

You are kindly invited to find out how you can benefit from our services.

Clear Words Translations
Córdoba, Argentina
info@clearwordstranslations.com
www.clearwordstranslations.com

Save Translation Cost with HyperSTE

HyperSTE is the leading quality assurance software for standardized documentation.

Benefits include
- Up to 30% in cost savings on translation and localization
- Up to 40% in reduced word count
- Quality improvement in writing and translations
- Quality assurance and quality measurement for content
- Up to 30% in reduced product cycle time
- Up to 40% reduction in overall documentation cost
- Improved safety and customer service
- Facilitates DITA, ST1000D, SCORM, CMS and XML

Tedopres International, Inc.
Austin, Texas USA
ste@tedopres.com • www.simplifiedenglish.net

TermNet – International Network for Terminology

TermNet, the International Network for Terminology, is a forum for companies, associations and universities that engage in terminology. Terminology is considered and promoted by TermNet as an integral and quality assuring part of any product and service in the areas of
- information and communication
- classification and categorization
- translation and localization

If you would like to join the international community, please visit www.termnet.org and contribute to our blog.

TermNet – International Network for Terminology
Vienna, Austria
termnet@termnet.org • www.termnet.org

Translating Services into 70 Languages

We provide translation services into 70 languages using the most modern technology for clients throughout the whole world.

- CEET provides translation, proofreading, localization, DTP, interpreting, voice-over and cultural consulting in all major world languages with special expertise in Central and Eastern European languages.
- We approach all projects with respect to customers’ needs and the cultural uniqueness of each country because we believe the language of your firm communicates who you are to the audience.

SDL
Berkshire, UK
www.lspzone.com

CEET Ltd.
Prague, Czech Republic
info@ceet.eu • www.ceet.eu

Translation Services into 70 Languages

We provide translation services into 70 languages using the most modern technology for clients throughout the whole world.

- CEET provides translation, proofreading, localization, DTP, interpreting, voice-over and cultural consulting in all major world languages with special expertise in Central and Eastern European languages.
- We approach all projects with respect to customers’ needs and the cultural uniqueness of each country because we believe the language of your firm communicates who you are to the audience.

SDL
Berkshire, UK
www.lspzone.com

CEET Ltd.
Prague, Czech Republic
info@ceet.eu • www.ceet.eu

Translation Services into 70 Languages

We provide translation services into 70 languages using the most modern technology for clients throughout the whole world.

- CEET provides translation, proofreading, localization, DTP, interpreting, voice-over and cultural consulting in all major world languages with special expertise in Central and Eastern European languages.
- We approach all projects with respect to customers’ needs and the cultural uniqueness of each country because we believe the language of your firm communicates who you are to the audience.

SDL
Berkshire, UK
www.lspzone.com

CEET Ltd.
Prague, Czech Republic
info@ceet.eu • www.ceet.eu

Translation Services into 70 Languages

We provide translation services into 70 languages using the most modern technology for clients throughout the whole world.

- CEET provides translation, proofreading, localization, DTP, interpreting, voice-over and cultural consulting in all major world languages with special expertise in Central and Eastern European languages.
- We approach all projects with respect to customers’ needs and the cultural uniqueness of each country because we believe the language of your firm communicates who you are to the audience.

SDL
Berkshire, UK
www.lspzone.com

CEET Ltd.
Prague, Czech Republic
info@ceet.eu • www.ceet.eu
BRIDGING THE GAP BETWEEN AUTHOR AND TRANSLATOR

Combined with MadCap’s industry-leading authoring and multimedia applications, MadCap offers the most powerful integrated authoring and localization workflow available, allowing for maximum translation reuse, reduced project cycles and costs, and vastly improving time to market.

- Streamline and manage the entire translation process regardless of the TM tool being used
- New Project Packager takes the guesswork out of what needs to be translated. Receive and send all necessary translation candidates (including multimedia elements) in a single ZIP
- Statistical reports show detailed information on translation status including what has/has not been translated, how many words/segments translated, etc.
- Seamless import of content created in Flare, DITA, Microsoft Word, RTF, TXT, XHTML, HTML and more
- Built-in translation memory and full TMX support
- Integrated terminology database
- And much more...

Download your free 30-day trials NOW!
www.madcapsoftware.com | +1 (858) 320-0387
Tthis guide is a component of the magazine *MultiLingual*. The ever-growing easy international access to information, services and goods underscores the importance of language and culture awareness. What issues are involved in reaching an international audience? Are there technologies to help? Who provides services in this area? Where do I start?

Savvy people in today’s world use *MultiLingual* to answer these questions and to help them discover what other questions they should be asking.

*MultiLingual*’s eight issues a year are filled with news, technical developments and language information for people who are interested in the role of language, technology and translation in our twenty-first-century world. A ninth issue, the Resource Directory and Index, provides listings of companies in the language industry and an index to the previous year’s content.

Two issues each year include *Getting Started Guides* such as this one, which are primers for moving into new territories both geographically and professionally.

The magazine itself covers a multitude of topics.

**Translation**

How are translation tools changing the art and science of communicating ideas and information between speakers of different languages? Translators are vital to the development of international and localized software. Those who specialize in technical documents, such as manuals for computer hardware and software, industrial equipment and medical products, use sophisticated tools along with professional expertise to translate complex text clearly and precisely. Translators and people who use translation services track new developments through articles and news items in *MultiLingual*.

**Language technology**

From multiple keyboard layouts and input methods to Unicode-enabled operating systems, language-specific encodings, systems that recognize your handwriting or your speech in any language — language technology is changing day by day. And this technology is also changing the way in which people communicate on a personal level — changing the requirements for international software and changing how business is done all over the world.

*MultiLingual* is your source for the best information and insight into these developments and how they will affect you and your business.

**Global web**

Every website is a global website, and even a site designed for one country may require several languages to be effective. Experienced web professionals explain how to create a site that works for users everywhere, how to attract those users to your site and how to keep the site current. Whether you use the internet and worldwide web for e-mail, for purchasing services, for promoting your business or for conducting fully international e-commerce, you’ll benefit from the information and ideas in each issue of *MultiLingual*.

**Managing content**

How do you track all the words and the changes that occur in a multilingual website? How do you know who’s doing what and where? How do you respond to customers and vendors in a prompt manner and in their own languages? The growing and changing field of content management and global management systems (CMS and GMS), customer relations management (CRM) and other management disciplines is increasingly important as systems become more complex. Leaders in the development of these systems explain how they work and how they work together.

**Internationalization**

Making software ready for the international market requires more than just a good idea. How does an international developer prepare a product for multiple locales? Will the pictures and colors you select for a user interface in France be suitable for users in Brazil? Elements such as date and currency formats sound like simple components, but developers who ignore the many international variants find that their products may be unusable. You’ll find sound ideas and practical help in every issue.

**Localization**

How can you make your product look and feel as if it were built in another country for users of that language and culture? How do you choose a localization service vendor? Developers and localizers offer their ideas and relate their experiences with practical advice that will save you time and money in your localization projects.

**And there’s much more**

Authors with in-depth knowledge summarize changes in the language industry and explain its financial side, describe the challenges of computing in various languages, explain and update encoding schemes, and evaluate software and systems. Other articles focus on particular countries or regions; specific languages; translation and localization training programs; the uses of language technology in specific industries — a wide array of current topics from the world of multilingual computing.

If you are interested in reaching an international audience in the best way possible, you need to read *MultiLingual*. 

---

Subscribe to *MultiLingual* at www.multilingual.com/subscribe

---

October/November 2009 • www.multilingual.com/gsg