



MultiLingual

WRITING FOR TRANSLATION

October/November 2009

GETTING STARTED: **Guide**



**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**

OPTIMIZED MT FOR HIGHER TRANSLATION QUALITY

LORI THICKE

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 solved an IT problem using Microsoft's automatically translated knowledge base.

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

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 10-20. The source text may or may not be consulted at the same time.



The current localization paradigm.

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

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 post-editor. 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 simship. 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

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 of 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.

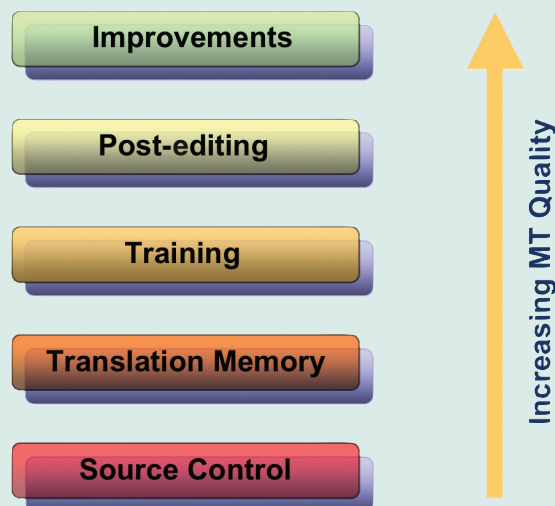
Microsoft concluded at a presentation to the 11th Machine Translation Summit in Copenhagen, Denmark, in September 2007 that “customer satisfaction numbers for machine translated articles is comparable to and sometimes exceeds original English!”

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.



Five factors influence increasing MT quality.

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

However — and this is a big *however* — the data must be good, clean data, following the garbage-in-garbage-out truism. As Microsoft says, “we can never have enough clean, parallel data.” And it must be domain and client-specific data: no point training the system on EU corpora if you’re a car manufacturer.

In rule-based machine translation (RBMT), this training is even more specific, involving data mining to create domain-specific dictionaries created for terminology entries including “Do Not Translates” and graphic user interface strings. This expert training of the engine creates the grammatically coded glossaries that will do the work of imprinting in-house terminology on the system. This is actually trickier than it sounds and requires a linguist trained in MT’s idiosyncrasies to avoid inadvertently creating errors and making the output worse, rather than better. This can occur when terms are coded incorrectly (a verb as a noun) but also when coding correctly but failing to take into consideration how the system will react to exceptions. If training the engine is the *sine qua non* of quality MT, it is also one of the greatest barriers because relatively few linguists know how to correctly tune MT systems, and few resources exist to tell them how to do it.

Upstream of the actual MT processing, another activity is important to optimizing MT output: controlled authoring, or language control of the source content. Long, convoluted sentences do not lend themselves to MT, no matter how well trained the system is. Authoring guidelines specify, for example, that technical writers use short, simple, declarative sentences, employ the active and not the passive voice, avoid parenthetical expressions in the middle of a sentence and so on. And while we humans may understand text that is rife with grammatical errors, no MT system will.

Where the source text already exists or where in-house documentation teams are resistant to applying the principles of controlled language for authoring, there is another solution. Using automatic normalization or running source text through a QA program may bring a noticeable improvement to the ability of your MT engine to understand and translate your text.

TM leveraging is another step in MT optimization. Even a well-trained MT engine is no replacement for the human translations contained in TMs, assuming they’re of high quality. It’s important therefore to develop the processes that will increase TM leveraging.

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 courseware 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. **G**