

# Improving MT results: a study

*Lori Thicke*

If machine translation (MT) has gone mainstream, our guess is that this has more to do with changed expectations than with improved technology. That MT technology has advanced goes without saying, but the biggest change may be that users no longer expect high-quality translations “out of the box.” Most users anticipate having to invest in some customization work to get the kind of MT output that is as close to human quality as possible.

With experience, we’ve found that MT gains start to become interesting when engine customizations are paired with other optimizations either before, after or during the MT process. Whether taking a rule-based (RBMT) or statistical (SMT) approach, there’s no doubt that a well-trained engine pays out the biggest dividends. However, once you have an optimized engine and an iterative process in place to improve it, what other ways are there to get better MT results?

This is the question we asked ourselves at Lexcelera. Specifically, we wanted to know which optimizations would give us the biggest bang for our buck. While enhancements can go on forever, we wanted to identify which ones were most effective in improving the quality of the raw output, without adding a significant burden to the process. Optimizations of the MT process most commonly involve training the engine on the target terminology – also known as customizing. However, other places to improve results include the vitally important step of re-training the engine with the feedback from real projects as well as improving the source and correcting the target, preferably through some automated procedures.

While the best results spring from working on all fronts, for the purposes of our study, we decided to isolate just one: improving the source text in keeping with Global English guidelines. We conducted our study with an RBMT engine because a rule-based approach would be more sensitive to improvements in grammar. RBMT actually parses a sentence to understand it, so logically it would gain more from a linguistically improved source. However, in order to benefit from SMT’s sentence fluency, we chose the new SYSTRAN hybrid engine.

## The study

To measure the effectiveness of source text improvements, we decided to use post-editing productivity as our metric. Although other measures – such as the BLEU score – are helpful in comparing trained engines, we wanted a measure that correlated fully with human evaluations of quality as well as with speed and cost.

Though the debate rages regarding the best quality metric to apply, we find that post-editing productivity – that is, the average time it takes a post-editor to bring a translation up to a fully human standard – correlates best with other measures of quality such as the LISA QA model, not to mention the admittedly subjective judgments of humans. Furthermore, given that the quality of the raw MT output determines the speed at which a post-editor can progress and thus determines a customer’s cost

*Lori Thicke is cofounder and general manager of Lexcelera, cofounder of Translators without Borders, and a member of the MultiLingual editorial board.*



Source (unedited and edited)	Target (raw MT output)
Unedited: Understanding the differences between owned and checked out alerts is critical to understanding SAS® Anti-Money Laundering.	La compréhension des différences entre les alertes possédées et Extraites est critique au SAS® Anti-Money Laundering de compréhension.
Edited: <b>In order to</b> understand SAS® Anti-Money Laundering, <b>you need to</b> understand the differences between owned alerts and checked out alerts.	Afin de comprendre le SAS® Anti-Money Laundering, vous devez comprendre les différences entre les alertes détenues par un autre utilisateur et les alertes bloquées.
Note that with the improved source text, the post-editor needed to make only one modification to the raw MT output: Afin de comprendre le <b>fonctionnement</b> de SAS® Anti-Money Laundering, vous devez comprendre les différences entre les alertes détenues par un autre utilisateur et les alertes bloquées.	

Rule 1: Use active verbs and avoid the gerund.

Source (unedited and edited)	Target (raw MT output)
Unedited: Risk-factor-only alerts can be identified by the Scenario and Triggering Values columns on an alert list window.	Des alertes de type facteur de risque uniquement peuvent être identifiées par le scénario et des colonnes Valeurs de déclenchement sur une fenêtre de listes des alertes.
Edited: <b>To identify a</b> risk-factor-only alert, the Scenario column of the alert list window displays either ML_Risk or TF_Risk.	Pour identifier une alerte de type facteur de risque uniquement, la colonne Scénario de la fenêtre de listes des alertes montre ML_Risk ou TF_Risk.
Again, the improved text requires little modification at the post-editing stage: Pour identifier une alerte de type facteur de risque uniquement, la colonne Scénario de la fenêtre de listes des alertes <b>indique</b> ML_Risk ou TF_Risk.	

Rule 2: Avoid the passive voice.

Source (unedited and edited)	Target (raw MT output)
Unedited: Alerts are displayed on alert list windows, which provide tools and information to aid users as they determine whether alerts represent suspicious activity that should be reported to authorities.	Des alertes sont montrées sur les fenêtres de listes des alertes, qui fournissent des outils et des informations aux utilisateurs d'aide pendant qu'elles déterminent si les alertes représentent l'activité suspecte qui devrait être rapportée aux autorités.
Edited: Alerts are displayed <b>in</b> alert list windows. <b>The alert list windows</b> provide tools and information <b>that help</b> users determine whether alerts indicate suspicious activity that should be reported to authorities.	Des alertes sont montrées dans des fenêtres de listes des alertes. Les fenêtres de listes des alertes fournissent les outils et les informations qui aident des utilisateurs à déterminer si les alertes indiquent l'activité suspecte qui devrait être rapportée aux autorités.
This is actually two rules in one. Both shorter sentences in general and limiting the text to one idea per sentence yield better MT results. The post-editor made these changes: <b>Les alertes s'affichent</b> dans des fenêtres de listes des alertes. Les fenêtres de listes des alertes fournissent les outils et les informations qui aident des utilisateurs à déterminer si les alertes indiquent une activité suspecte qui devrait être signalée aux autorités.	

Rule 3: Use short sentences with just one idea.

savings, post-editing productivity provides valuable information about quality, speed and cost.

Using post-editing productivity as our metric, we set out to measure the impact of improving the source text on the quality of RBMT output and thus on the speed of post-editing. Content was provided by SAS Institute, the largest independent vendor of business intelligence software. The study collaborators were John Kohl, technical editor/linguistic engineer at SAS Institute and author of *The Global English Style Guide* (2008), and Richard Meneglier, localization project manager in Lexcelera's Paris office.

The test document was an 880-word, three-topic portion of the online Help for SAS Anti-Money Laundering Software. This document was chosen because it was very well written according to the standards that most companies follow, but it was not written with translation in mind. It contained no grammatical, spelling or terminology errors, but it violated a number of the Global English guidelines described in Kohl's style guide that are known to have an effect on the quality of the output produced by RBMT systems. Although the document consisted of Help topics, the topics that were selected presented conceptual information; they were not task-oriented instructions. Task-oriented instructions would likely have been simpler syntactically, presenting fewer opportunities for making the information more suitable for RBMT.

The SAS European Localization Center provided translations for about 500 technical terms and user-interface labels that occur in SAS Anti-Money Laundering documentation. As technical editor, Kohl determined that 56 of those terms occurred in the test document, and Meneglier, the project manager, coded those terms and used them as a "mini-training" of the SYS-TRAN hybrid engine. The technical editor then edited the source text according to Global English rules. This gave us two versions of the source document: edited and unedited. To compare the results of this pre-editing with the results of engine training, we tested both the edited and unedited source text using both a trained and untrained MT engine. This meant that we were actually testing four scenarios: untrained MT engine with an unedited source document; untrained MT engine with an edited source document; trained MT engine, unedited source document;

and trained MT engine, edited source document. Each file was post-edited separately, and the post-editing time was thoroughly tracked.

## Results

The untrained engine underperformed with both unedited and edited source. Not surprisingly, the worst versions came from the MT “out of the box” with no engine training. The system struggled to understand the basic terminology, with the result that the post-editor had to spend more time fixing terms. Additionally, the unedited text, breaking the rules of Global English, was more difficult for the machine to understand, just as it would have been for a human reader.

Interestingly, in the absence of a correctly trained engine, even well-authored text didn't fare noticeably better. With an untrained engine and unedited source, the post-editing productivity was 5,587 words per day, a decent rate considering that the average human translation rate is 2,500 words. But that rate is below potential as far as MT goes. With an untrained engine and an edited source document, this rate was only slightly better: 6,208 words per day.

However, the trained engine reached peak performance, particularly with edited source material. Although measuring the impact of engine customization wasn't the purpose of this study, it was abundantly clear that this yields the most significant gains. Once the dictionaries were added to customize the engine, the output quality improved dramatically, regardless of whether the source text was edited or not. With a trained MT engine, post-editing productivity increased to 7,880 words per day, even on the unoptimized source content. This reflected a significant improvement in the output quality, mainly due to the inclusion of the appropriate terminology in the engine, which avoided excessive terminology look-up, which was the largest time sink of all post-editing activities. However, with the unedited source, grammatical mistakes still remained in the output, and the resulting post-editing productivity was lower than it could have been.

Not surprisingly, the best combination of activities for increasing post-editing productivity was to have a trained engine and optimized (in this case, pre-edited) source content. With



## More Than Quality – With Efficiency!

We offer fast, reliable translations for global players. With two decades of experience we are a leading international company for high-quality localization into German and traditional translation services into most Western European languages. Our team's expertise lies in the fields of software, hardware, medical engineering, finance, technology, legal and marketing. Tight deadlines, large volumes or projects requiring technical specialists are our core competence. Rheinschrift is EN 15038:2006 certified.

### Rheinschrift Übersetzungen, Ursula Steigerwald

Cologne, Germany  
contact@rheinschrift.de • www.rheinschrift.de



## Europe's No. 1 Greek Localizer

Since 1986, EuroGreek has been providing high-quality, turnkey solutions, encompassing a whole range of client needs, for the following language combinations:

- English into Greek
- Greek into English
- German into Greek
- French into Greek

All EuroGreek's work is produced in our Athens production center and covers most subjects:

- Technical
- Medical/Pharmaceutical
- IT/Telecommunications
- Economics/Legal

All EuroGreek's work is fully guaranteed for quality and on-time delivery.

### EuroGreek Translations Limited

London, UK • Athens, Greece  
production@eurogreek.gr • www.eurogreek.com



## Your Polish Competence Center

Since 2000, Ryszard Jarza Translations has been providing specialized Polish translation, localization, marketing copy adaptation, and DTP services. We focus primarily on life sciences, IT, automotive, refrigeration and other technology sectors.

We have built a brilliant in-house team made up of experienced linguists and engineers who guarantee a high standard of quality while maintaining flexibility, responsiveness and accountability. Our services are certified to EN 15038:2006.

### Ryszard Jarza Translations

Wrocław, Poland  
info@jarza.com.pl  
www.jarza.com.pl



MEDILINGUA  
TRANSLATIONS

## Medical Translations

MediLingua is one of Europe's few companies specializing in medical translation. We provide all European languages (36 today and counting) and Japanese as well as the usual translation-related services. Our 250-plus translators have a combined medical and language background.

We work for manufacturers of medical devices, instruments, in-vitro diagnostics and software; pharmaceutical companies; medical publishers; national and international medical organizations; and medical journals.

Call or e-mail Simon Andriesen or visit our website for more information.

### MediLingua BV

Leiden, The Netherlands  
simon.andriesen@medilingua.com  
www.medilingua.com

## Potential for using controlled authoring software to facilitate editing

Pre-editing the source content for this study required a time investment of 32 minutes, which would not have been necessary if the author had followed the Global English guidelines to begin with. Automation of this process is, of course, more logical. In fact, one of the goals of the project was to get an idea of how useful controlled authoring software would be for drawing the author's or editor's attention to issues that are problematic for rule-based machine translation (RBMT).

The technical editor had considerable experience supporting and customizing a controlled authoring application — acrolinx IQ software — for SAS. Because SAS is not yet using MT, the grammar rules, style rules and terminology restrictions that have been implemented with acrolinx IQ at SAS focus on issues that are problematic for human translators or for nonnative speakers. Many of those issues are also problematic for MT, but many other issues are problematic only for MT.

This project made it clear that the process of editing for MT could be facilitated by extending the acrolinx IQ rule set that is used at SAS. For example, an acrolinx IQ rule called `repeat_head_noun` flags coordinated constructions and suggests repeating the head noun. In the following example, the article *the* must also be repeated:

Not optimized for MT:

*The color and label customizations will be changed.*

Optimized for MT:

*The color customizations and the label customizations will be changed.*

The optimized version is much more suitable for RBMT. For stylistic reasons, most companies would not choose to implement that rule if they are using only human translation, but the rule is well worth implementing if RBMT is being used. It would also be easy to develop an acrolinx IQ rule that would flag all instances

of pronouns such as *it*, *they*, *yours* and *one's*, suggesting that they be replaced by noun phrases.

Not optimized for MT:

*Click on the user's name to remove the alert from that user's My Alerts list and to place it in yours.*

Optimized for MT:

*Click on the user's name to remove the alert from that user's My Alerts list and to place the alert in your list.*

There are many other MT-specific issues that controlled authoring software cannot target as reliably. For example, SAS uses an acrolinx IQ rule that flags all instances of future tense. Because future tense is often appropriate and necessary, most of the "errors" that the rule flags are false alarms. Nevertheless, it is useful to draw the author's or editor's attention to each instance and to ask them to change the tense to present wherever possible. This type of rule is referred to at SAS as a verification-and-assessment rule.

Companies that use controlled authoring software need to consider how many verification-and-assessment rules their authors or editors will tolerate. If there are too many of these rules, then authors or editors will encounter a high percentage of false alarms and might refuse to use the controlled authoring software. At SAS, authors use a smaller set of rules that excludes most of the verification-and-assessment rules. Editors use the complete set and are more tolerant of false alarms for two reasons: They hold final responsibility for the quality of documentation deliverables, and they are already in the habit of verifying and assessing a huge number and range of issues in the course of their duties.

— John Kohl, technical editor/linguistic engineer at SAS Institute and author of *The Global English Style Guide*.

this combination, the output quality was very good, and many of the grammatical mistakes disappeared. The sentence structures in the source text were simplified, which enabled SYSTRAN to process the content. Post-editors were thus able to get away with just a small tweak here or there to bring the sentences up to fully human quality. The productivity was exceptional: 9,677 words per day. To summarize our results: untrained MT is two times faster to post-edit than to translate from scratch; trained MT is three times faster; and trained MT with source control is four times faster.

### High-impact source text improvements

Given the improved productivity with improved source content, we then moved to the second goal of our project, which was to identify which among the numerous rules of Global English have the most impact on MT quality. Upon analysis, we

identified Rules 1, 2 and 3, illustrated on page 38, as having a high impact. Each version of the source text, whether unedited or edited, is on the left, with the resulting MT into French on the right. Modifications are shown in red.

Of all the improvements that can be made to source text to improve its machine translatability, we would agree with Greg Oxtton of the Consortium for Service Innovation that the single most powerful rule for technical writers is to limit themselves to one idea per sentence. As an added benefit, text that is easier for an MT engine to understand is also easier for humans to understand.

In conclusion, a well-trained engine with source content that follows Global English guidelines generates the highest MT quality. Our starting point for this study was an untrained SYSTRAN hybrid engine, and even out of the box the output was twice as fast to post-edit than to translate from scratch. However, simply customizing

the engine with the correct terminology resulted in a post-editing processing speed that was three times faster than a fully human translation. Adding well-authored text to the mix resulted in a post-editing productivity four times that of a traditional translation. This promising finding points to the gains that are possible when using any type of source control whether a result of controlled authoring using a program such as acrolinx IQ (see John Kohl's sidebar above); pre-editing in a manual or automatic process, including text normalization; or respecting just a very few high-impact guidelines such as sentences that reflect just one idea.

Furthermore, an RBMT engine seems particularly sensitive to improvements in grammatical structure. With a trained RBMT engine and a good source text, the result is measurably higher quality MT output, which means increased post-editing speeds and decreased localization costs. **M**