

# AMTA-2010: Content Quality for Better MT: A Practical Guide to Quality at the Source

**Jennifer Beaupre**  
Acrolinx  
1160 Ithaca Dr  
Boulder, CO 80305  
Jennifer.Beaupre@acrolinx.com

**Kent Taylor**  
Acrolinx  
1160 Ithaca Dr  
Boulder, CO 80305  
Kent.Taylor@acrolinx.com

## Abstract

With pressure to offer content in many languages, many companies are considering machine translation for faster delivery and lower translation costs, yet MT is notorious for poor quality translation. How can you improve your content quality to make MT work for you?

High quality source content eliminates many of the common roadblocks for using machine translation effectively. In this presentation, Jennifer Beaupre, Marketing Director and Kent Taylor, GM, acrolinx, will review what best practices have taught us about these topics:

- 1 Why is source content important when using machine translation?**
  - 2 How does source content affect translation costs?**
  - 3 How can source content improve the quality of MT output?**
- 3.1 Electronically-available resources**

**Industry Trend: Quality at the Source**  
*Understanding how an Information Quality Strategy is essential in today's business world.*

In current times, companies must reach global markets faster if they want to remain competitive. To most effectively reach those markets, corporate content such as websites, product information, and sales materials must be translated into their customers' native language. However, translation and localization is often a roadblock to successfully releasing products into international markets quickly. The two biggest hurdles for localization are cost and time. Traditional translation done with a translation vendor can range in cost from 0.15 to 0.30 cents per word depending on the provider. That can mean millions of dollars when translating millions of words in multiple languages. Additionally, localization is often the last step for a product release. Once a product is created, the release is delayed until all of the translation is completed. Every day spent waiting on translation is a day that companies are losing money and market share. Imagine if the BlackBerry™ smart phone was released in France two months before the iPhone™ because of translation delays. How many potential customers would have been lost? Could Apple ever get those customers back?

This article will focus on strategies for dealing with the biggest challenges with localization- cost, time, and quality. You will learn how to create an Information Quality strategy to decrease the cost of translation and contribute to your company's strategic objectives of reaching global markets faster.

## What is Information Quality?

It is important to understand what Information Quality means and how to measure it to create

a strategy for reducing localization costs and addressing international release delays. Information Quality refers to the measurement of excellence for a company's written content. Corporate content in multiple formats (XML, HTML, Word documents, Power Point slides, etc) can be measured for accuracy, readability, clarity, consistency, conformance to corporate and industry standards, and translatability. This measurement can be used as an Information Quality standard. Software products like acrolinx IQ™ can provide data on the quality of corporate information and help improve this quality over time.

Many companies look downstream to their translation vendors to solve their cost and time challenges. Yet, looking at the content at the beginning of the Information lifecycle adds more value to the organization. By improving the source content, companies drive down the supply chain costs, improve their customer experience and accelerate time-to-market at the same time. Additionally, they can improve the results of machine translation by putting higher quality information in at the beginning, which in turn produces a better result from the machine.

### ***Writing to improve Information Quality***

Information developers are an essential part of an organization. Yet, their work is often undervalued. The source content that they are creating will have an effect on customer satisfaction, translation costs, and publishing costs. Why is their work so important? It is the information that your customers are seeing. Customers want to be spoken to in a language that they understand! That means that they want to read and buy in their native language, feel comfortable with the product or service, and understand the information that they are reading. If, for example, an American travels to Africa and books a room at a Hilton hotel, he or she expects it to have the same look, feel, and service as a Hilton hotel in the United States. Customers feel the same way when they are reading information in their native language. They expect it to have the same high quality as the English content. The rule of thumb is: when communicating brand, translating into different languages, or indexing information for searchability, the better the source content, the better the output.

You can improve source quality in many ways. First, consider the audience. Do not reference information that is not relevant in a global way. Some good examples are: holidays, date and time formats, and catchy phrases that wouldn't translate well into another language. Second, define your corporate terminology. Using consistent terms for your company's products services, and messages affects brand recognition, customer satisfaction and the cost of translation.

Developing a corporate style guide and terminology data base, making them available for everyone in the organization, and using technology to alert writers to improper terminology, will help ensure consistency throughout your content. Additionally, decreasing the amount of content by eliminating unnecessary words and phrases, writing the same information the same way, and reusing previously written and/or translated content will have a dramatic effect on cost and time concerns.

### **How does this help in Machine Translation?**

Machine Translation engines have a reputation of producing poor quality translations that generally cannot be used for customer facing content. However, when companies implement an information quality strategy from the beginning of their content lifecycle, the results can improve dramatically. When content developers take the following considerations into account, the result of machine translation can be improved.

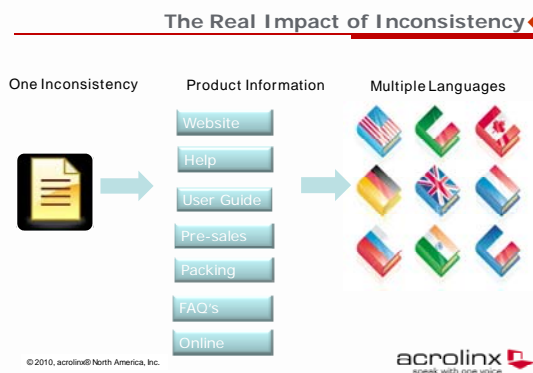
#### ***Word volume***

Machine translation engines will produce better results if the volume of words to be translated is decreased. A rule of thumb is: the simpler the text going in, the better the translation. We have all heard the term garbage in, garbage out-this is true of both machine and human translation. Additionally, sentence length is important when thinking about improving results. The longer the sentence, the harder it is for the machine to produce a comprehensible translation. It is key to cut out irrelevant or non-essential words to simplify as much as possible.

#### ***Inconsistencies and Errors***

Many times, translation quality decreases and costs skyrocket because of errors, updates and inconsistent language in the source content. When

an error or update is introduced and fixed at the source, it is a minimal cost to make the change. If that error is then introduced into multiple channels for product information such as website, user guides and online help, that cost is multiplied by each form of content. When that error is then translated into multiple languages, the cost is exponentially greater. Mistakes and inconsistencies in the source content can proliferate throughout all corporate information. The cost to fix it at the beginning of the content lifecycle is significantly less than finding it after translation. Additionally, machine translation results will be of higher quality in every language if the errors are fixed at the source. *The key to keeping costs down and high quality output is to insist on quality source content that is consistent, streamlined, and accurate. Information Quality directly affects the downstream cost and quality of translation.*



### How does this help me get products to market faster?

Simultaneous shipping, or releasing products in all markets and languages at the same time, seems unachievable for many product manufacturers. Typically, companies can't afford to release a product while they wait for all of their product information to be localized. The localization process simply takes too long and waiting would cost too much in revenue and market share. High quality information at the source helps get products through translation and to target markets faster in several ways. First, it decreases the volume of content to translate which decreases the time it takes to translate it. Second, it decreases the time that content is in the review cycle because high quality information is easier to understand and translate. Next, it prevents errors and inconsistencies in the content that cause delays in the release.

Finally, it improves the quality of the information so that it is possible to produce high quality translations in a machine environment which can be done both faster and cheaper.

### An Information Quality Strategy should:

- **Eliminate unnecessary content and encourage reuse**
- **Enforce corporate terminology and style guides**
- **Make it easy for authors to recognize where they can make improvements**
- **Provide Information Quality metrics**

### Benefits:

- **Improve the results of MT**
- **Decrease translation and publishing costs**
- **Get to market faster**
- **Improve efficiency for authors**
- **Easier to understand**

### Summary

Implementing Information Quality standards across the enterprise can benefit organizations in multiple ways. Mostly, they provide benefits in cost, time and efficiency. Bottom line: Without an Information Quality strategy that empowers Information Developers to maintain quality in the source content, companies will lose time and money. Also, it is essential to measure and identify quality trends. Companies need to put in place quality metrics tools and provide objective, real-time feedback to the writers in order to maintain quality standards. Whether companies want to measure and enforce quality manually or rely on an Information Quality Management tool such as acrolinx IQ™, it is apparent that quality information needs to part of every company's strategic mission.