# Introduction to the SAE J2450 Translation Quality Metric

by Rick Woyde

One of the daily challenges that all translation companies face is producing and delivering high-quality defect-free translations on a regular basis for a variety of customers, with different needs and expectations. From my perspective, we appear to really be in the business of providing custom multilingual publishing services to each and every client. While deciding on what services to provide and on which customers to ultimately focus and serve

Rick Woyde

we also have to produce acceptable levels of translation quality along the way. This alone is a huge challenge, and even more so for multi-language vendors that provide translation services in a variety of different languages for a variety of different subject matters while managing these activities across the globe. To help us manage our businesses we have a number of tools at our disposal. To measure financial success we have income statements, balance sheet statements and cash flow statements. All of

them tell the same story from different perspectives. And those are just a few of the financial tools available. Likewise, we

> **Others believe that almost everything can be measured, including translation. Certainly some things seem more difficult, if not impossible, to measure. Can you measure art or photography?**

can measure our sales and marketing success by capturing sales data, such as the Number of Quotes produced, Sales Closed and Delivered, Work in Progress, and others. We can also measure productivity, by the number of words translated per day, actual time compared to estimated time and, possibly most importantly, on-time delivery. All of theses activities are measurable.

## Why can't we measure translation quality?

By now you might have an idea where we are headed. However, when the subject of measuring translation quality comes up no one seems to agree. Many people believe that translation by its nature, being subjective, cannot be measured. Others believe that almost everything can be measured, including translation. Certainly some things seem more difficult, if not impossible, to measure. Can you measure art or photography? For that matter, how do you measure Olympic figure skating? Those activities seem extremely subjective to me. And I agree that translation shares similar qualities. There is subjectiveness to translation that can't be measured. Each company devises a strategy of its own to accommodate each customer. However, I do think there are some aspects of translation that can be easily and accurately measured. For instance, spelling errors come to mind and I think we can also agree on when a wrong term is used. So, who's right? In my opinion both camps are. In my experience, there are key translation qualities that can be measured and should be measured.

There are also other issues, mainly stylistic, that cannot be measured.

## So who is the SAE and what is J2450?

SAE is the acronym for the Society of Automotive Engineers. You'll most likely hear them refer to themselves simply as the SAE and they serve more than just the automotive community. Founded in 1905, the SAE is one of the oldest standards creating organizations in the United States.

> **Since its inception, our committee members have included representatives from GM, Ford and Daimler Chrysler, and translation suppliers to the automotive industry.**

The SAE is a resource for technical information and expertise used in the design, building, maintenance, and operation of self-propelled vehicles for use on land or sea, in air or space. Additional information about the SAE can be found at www.sae.org. I first learned about the J2450 initiative back in January 1999 from Kurt Godden, then Manager of Translation Process Development at General Motors North America. Kurt was involved in various translation activities for GM at the time. His responsibilities included researching language technologies including translation memory, machine translation and database driven publishing environments. Kurt's background is in computational linguistics and he had been working for GM for at least fifteen years by the time I met him. So, he had been in a manufacturing environment for quite some time. As you may or may not know, manufacturers are big believers in measuring productivity and quality and have been doing so for many years. After all, they are in the business of building complex cars within the smallest of tolerances and only after making billion dollar investments upfront. So, Kurt set up a task force, (which I should mention is a subcommittee to the SAE Vehicle E/E Systems Diagnostic Committee), with the goal of creating a Translation Quality Metric to measure the translation quality of automotive service information. Since its

inception, our committee members have included representatives from GM, Ford and Daimler Chrysler, and translation suppliers to the automotive industry. It should also be mentioned that this was not the first time that a metric for measuring translation quality specifically for automotive materials had been suggested. In fact, one already existed. Under the direction of Oleg Kuzin, GM of Canada had already been using a translation quality metric prior to J2450. Shortly after arriving at GM, Oleg Kuzin met Pete Peters, a GM Quality Consultant, who organized in Oshawa a yearly Continuous Improvement Symposium for GM suppliers. Case studies using statistical tools were featured during the Symposium which all dealt with production problem resolution in GM plants. After attending a number of these meetings, Oleg realized that some of the quality problems encountered in industry were similar to those found in translation. With the help of Pete Peters, he devised a statistical tool that quantified the errors in an objective way and which could be used as an improvement tool. The GMCL SPC Translation Tool was born. It was tested for repeatability and reproducibility for a number of years and passed with flying colors. It was formally presented to the GM supplier community at a Continuous Improvement Symposium held in Oshawa in December 1997. It is this system that J2450 is trying to build upon.

Our task force began meeting in November 1998, first at different supplier offices and then at the SAE offices in Troy, Michigan. In 1999 and for most of 2000 we met once a month, usually on a Wednesday in the middle of the month. Since August of 2000 we've been meeting

> **As a quality metric, J2450 and its use should not be confused with a translation editing process.**

every other month. We've also held additional meetings in conjunction with the Annual SAE TOPTEC. TOPTEC is a Multilingual Documentation Symposium specifically for the Automotive Industry. It was first held in 1999 in Amsterdam and last year it was held in Dearborn, Michigan near the Ford Motor Company World Headquarters.

### How does the Metric work?

While the metric itself is a Score Sheet that enables evaluators to capture error types and quantities of translation errors, it was decided that additional documentation would be needed. To support the implementation and use of the metric in a real world environment our task force developed reference materials that include a reference document with guidelines for evaluators to follow, detailed definitions of all error categories, examples of many of the error types and a Quick Reference Guide. The Metric Score Sheet consists of four elements: the seven primary error categories, two secondary subcategories (i.e. serious and minor), two Meta rules to help

## Numeric Weights

| | Serious | Minor |
|---|---|---|
| Wrong Term | 5 | 2 |
| Syntactic Errors | 4 | 2 |
| Omissions | 4 | 2 |
| Word Structure and Agreement Errors | 4 | 2 |
| Misspellings | 3 | 1 |
| Punctuation Errors | 3 | 1 |
| Miscellaneous | 3 | 1 |

*Table 1: Measured errors*

decide ambiguities on the assignment of an error to the categories and sub-categories and, finally, the numeric weights for each category and secondary subcategory. The Metric measures six translation and grammar errors and allows for other error types that cannot be clearly attributable to be classified as miscellaneous errors, see Table 1.

As a quality metric, J2450 and its use should not be confused with a translation editing process. Quality assessments may influence translation-editing processes, but they do not replace them. The J2450 metric should only be regarded as one element in a total quality assurance process. The actual application of the metric is rather easy and non-cumbersome. As I mentioned earlier, the metric is a Score Sheet and the objective is to accurately identify and record the number of translation error occurrences. When evaluating translation quality, each error found by the evaluator is marked in two ways. First, it is classified into one of the seven categories described above, e.g. 'wrong term'. After its primary category has been identified, the evaluator determines if it is a 'serious' or a 'minor' error, according to how severe the evaluator considers the error. Both the primary (category) and the secondary (serious/minor) classification are judgment calls by the evaluator. After determining

> **Our current challenges include recruiting additional participants worldwide, continuing to evaluate the performance of the metric in the real world and determining whether or not we should broaden the scope of the metric to include the translation of owner's literature.**

the correct primary category and secondary category each error is multiplied by its weight and totaled. This number is then divided by the total number of words measured creating a translation quality score (TQS). See Example Box.

So what does this overall weighted score mean? It was decided early on that customers might have different quality goals and tolerances. Therefore, it is expected

## Example Box

For example, to determine the Category Score for the Syntactic Error category with a major weight of 4 and a minor weight of 2, assuming 1 major syntactic error and 2 minor syntactic errors:

$$\underline{1} * \underline{4} \quad + \quad \underline{2} * \underline{2} \quad = \quad \underline{8}$$

To take that a step further and determine an overall Score (in a document with 300 source text words):

$$\underline{8} + \underline{10} + \underline{6} + \underline{5} + \underline{2} + \underline{4} + \underline{8} = \underline{43}$$
(Sum of Weighted Scores in all 7 categories)

$$\underline{43} \text{ (Sum of Weighted Scores)} \div \underline{300} \text{ (Number of Words in Source Text)}$$
$$= \underline{0.143} \quad \textit{(Overall TQS)}$$

that each company implementing the metric would work towards determining a score that would meet those objectives.

To help deal with ambiguities two Meta rules were created to assist evaluators. The first rule is that when an error is ambiguous, an evaluator should always choose the earliest primary category as they appear on the score sheet. In other words, the order of the error categories is important as well. The second rule is that when in doubt as to the seriousness of an error, you should always choose "serious" over "minor". While reasonably complete definitions are available to help evaluators identify the major categories with some consistency, there may yet be room for ambiguity and the final category selection may, in fact, be arbitrary. The metric also purposely avoids attempting to measure the stylistic quality of a translation. To date, the scope of our work has been limited to automotive service information. The target customers that use these materials are service technicians. The J2450 translation Quality Metric was designed to be used to tag errors that are linguistic in nature. Therefore, a translation that is free of any J2450 errors may still be unacceptable for other reasons. The metric should be considered as only one criterion that companies would use to determine their overall effectiveness in managing their translation activities.

It must be acknowledged that unavoidable ambiguities that influence metric results

still remain. There are also translation quality issues that cannot be measured that may be more important to overall quality than the translation errors measured by the metric. But I think it's fair to say that the metric does help reduce the error types that it actually measures, such as spelling errors and wrong terms. That can only help. I think it's safe to say that we all want to deliver translations free of spelling errors and wrong terms. And I think we may all agree that producing high quality translations on a consistent basis requires a combination of managing different activities and resources such as; technology, processes, quality initiatives, service and, most importantly, people. Today our task force continues to meet every other month. Our current challenges include recruiting additional participants worldwide, continuing to evaluate the performance of the metric in the real world and determining whether or not we should broaden the scope of the metric to include the translation of owner's literature.

*Rick Woyde is President and CEO of Detroit Translation Bureau based in Troy, Michigan.. He can be contacted at rickw@dtbonline.com.*