The Datafication of Translation  Peng Wang


Dynamic Quality in the Business World  Peng Wang

The translation industry has come to a new era of datafication. This is a trend no one can stop and it also means human beings are much more thoughtful and considerate about their past work and legacy data. By continuously feeding the machine with data and interacting with machine through post-editing and manual data adjustment and annotation, human beings are producing smarter machine-operated systems. This includes Machine Translation (MT) engines, terminology management platforms, translation memory tools, translation management software, and multilingual annotation systems.

Translation-related data can be used in many ways and some of its applications go beyond the translation industry itself. At the TAUS Roundtable in Washington D.C. and the annual conference in San Jose, we saw the following areas where translation data are widely used.
Translation-related data to support machine translation

Parallel data, monolingual data and post-editing data help train machine translation engines. Parallel data and target language data are very important for the machine to learn the alignment and translation rules and gain the naturalness of the translation in the target language. Except for SYSTRAN rule-based MT engine, most of current MT engines are based on statistical analysis of the data. Domain customization in machine translation becomes more and more important in the MT world.

From a developer’s perspective, it makes much more sense to take a top down approach, i.e., starting from a very general topic to get some results first and then move down to more specific domains and fine-tune these results. In-domain data help improve the quality of machine translation. As Chris Wendt, Microsoft’s Group Program Manager - Machine Translation Microsoft Research, presented at the TAUS conference, Microsoft Translator customers achieve an average of 10 BLEU points increase through domain customization. That is a huge increase in machine translation and it undoubtedly shows the power of customization.

Domain-customized dictionaries, bilingual parallel data and monolingual target language data could generate different translation results, if not necessarily better results. Some MT engines such as TILDE Translator allow users to upload, share and use domain-specific translation memory to customize machine translation results. In other words, users can customize the machine translation engine by feeding the machine with different data. We can also see a different relationship between human beings and machines in this trend. Human beings used to have almost no control on the machine translation process except inputting or uploading the source texts to the machine. The new building-your-own-MT model gives more power to the human users, who start to realize the importance of collecting such data as translation memory, termbases and monolingual data, and start to manipulate the machine-feeding parameters so as to control machine translation results to some degree.

Translation-related data to support business decision-making

Translation-related data can also be used for business intelligence and help marketers solve business problems. The business industry collects, uses and analyzes data in order to make more informed decisions. The panel discussion moderated by Sergio Pelino (Google) at the TAUS conference, represented opinions and experiences in this regard from the buyer side - Salvo Giammarresi (PayPal) and Andrea Siciliano (Google), on the supplier side - Doug Knoll
(Welocalize), and Henry Wang (UTH International), a company that purchases data as a commodity.

The panelists agreed that translation-related data are useful in business management and it helps provide insight for strategic decisions in business. By analyzing the data and examples, it can help prevent fraud, recommend items to shoppers, predict emergency room waiting times, and improve localization performance. Language service providers can also use the data to select the right translators at the right time and under the right conditions. The data can also be useful for marketing people to predict the right expectations from customers. In some cases, translation-related data can help users segment the content and classify language combinations. For example, often we find data will not work across languages evenly. Some data will work well in some languages, whereas in some harder languages—or in other words, tier 2 languages—the data will not work well.

Here again, domain is an important parameter. In-domain translation-related data could be more useful in the aforementioned areas because these data are more targeted and relevant to the purposes of data analytics.

Anything that generates value might have a cost. Then what is the cost of data? What data are valuable to the company? What is the pricing strategy for this special soft product? This area is wide open and needs more discussions.

**Translation-related data to support predictions**
Translation-related data can assist human beings to make predictions. As Smith Yewell, founder and CEO of Welocalize, presented at the TAUS Roundtable in Washington D.C., most of the quality and on-time delivery problems in the localization industry are discovered after the fact, and too few localization programs are linked to measurable business outcomes. Our industry needs to introduce predictive analytics engines, just as many other industries have done.
There are many areas data can help us make predictions. For example, we can analyze the translation patterns of a translator, and predict some quality issues related to this person. At the **TAUS Insider Innovation Excellence Awards**, Olga Beregovaya from Welocalize demonstrated a new tool they developed, StyleScorer, which can help automatize the process of stylistic analysis and stylistic predication. In addition, as we mentioned earlier, humans can use translation-related data to predict items that shoppers may like and recommend them to shoppers. Data can also be useful for marketing people to predict the right expectations from the customers. Integrated translation management systems similar to Uber will allow us to make predictions on translation needs, translators’ availability, the right domains for translators to work with, an individual translator’s rate and translation quality issues. These are all promising areas for us to continue to explore.

**Closing Remarks: Innovation through collaboration**

The language industry is undergoing a huge revolution. In this revolution, data collection and data analytics play a very important role. Translation-related data have been widely applied in many areas beyond the translation industry itself. Many disciplines are involved in this process, for example, translation, corpus linguistics, computational linguistics, business management and neurology. Interdisciplinary collaboration is inevitable and it will definitely help realize a smarter data application on the translation and localization market.

As an educator and researcher, I have seen quite a few applications of theoretical frameworks and research findings in the industry. But I also see great potential in this regard. On the one hand, there are many more research findings and theories that the industry have not fully utilized, for example, discourse analysis theories, comparable translation-driven corpora, and textual and contextual anticipation. On the other hand, the academia should listen to the market and customize their research to meet the market needs. Together the industry and academia can promote innovations and push the profession forward.
The concept of quality is relational. At the center of the judgment, people are evaluating the relationship between the product and its reference criteria. The earliest benchmarks for translation quality are two types of text(s), or linguistic data. One is the source text and the other is non-translated texts in the target language. In the former case, quality evaluators compare the translation product against its source text in order to determine its faithfulness, or adequacy level as stated in the TAUS Dynamic Quality Framework (DQF) Adequacy/fluency guidelines; and in the latter, evaluators compare the translation against non-translated texts in the target language in order to judge the naturalness of the translation, or in TAUS’s term, the “fluency” of the translation.

Are these two linguistic benchmarks enough for us to make a decision about translation quality? Certainly not! In the real
world of business, many other factors come into play, including the content type, translation purpose, technology, budget, translators and clients. These benchmarks are far more dynamic.

**New Benchmarks for Translation Quality**

In recent years, the translation industry has seen much more diversified content categories, purposes, domains, and technologies in translation. These factors are more dynamic and should be incorporated into our judgment of the translation quality. At the TAUS Annual Conference in San Jose, California, in October 2015, the panelists discussed the following new quality data:

**Content categories and different ties of translation quality**

The source text content can be classified into different categories based on the quality requirements and the market needs. For example, some types of source texts such as user-generated content and translation texts that only a few people read will have different requirements for translation quality from those texts that are key to the business brand and marketing strategies. Based on their budget and target market, translation buyers may ask translators to produce different levels of quality for different content types. In this scenario, translation product is exactly like other commercial products. Buyers decide on the level of quality and vendors produce the required level of quality. This raises an issue for translators, though. Translators used to treat translation as an art or craftsmanship. They feel they need to polish their translation as best as they can so that the target text can be as fluent and adequate as possible. However, in the model where translation is treated as a commercial product, translators need to refrain themselves from producing translation results that are too good. There is no one-size-fits-all criterion to differentiate the translation products that are too good from those that are good enough. It all depends on the client’s purpose and each case might be quite different.

**Productivity per vendor**

There is no absolute quality. All quality is related to certain time constraints, as any translation task is done within a certain amount of time. When we measure the quality of the final product, we need to consider the factor of time and measure the productivity of translators.

Steve Richardson from LDS Church pointed out that when they try to measure translators’ productivity through the TAUS Dynamic Quality Framework platform, some translators are concerned about whether they can do something else (for example, going to the bathroom) during translation, while the software
is counting the time. This raises a very interesting topic. If we measure the productivity and use a software application to monitor the time for translation, we need to revisit the concept of translation time. In the past, translators did research, looked up words in a dictionary and learned new concepts during the process of translation. Many of them did not differentiate the time of translating from the time of learning and researching. If translators use a quality assurance platform such as TAUS DQF, which differentiates these two types of activities, ultimately their working habits will change and they may also have a strong motivation to separate the time of learning from the time of translating.

**Quality related data per vendor**

Many companies started or are interested in collecting data on translation error patterns and productivity scores for a specific translator or vendor. This quality related data could help them find the right translators at the right time in the right conditions. It can also help vendors improve their translation abilities and deepen their translation skills.

**Quality related data per language**

In most cases, translation quality will vary across languages. The machine translation engine, for example, might work for some languages whereas fail in some other languages. In which areas did the MT engine fail in these language combinations? Is it because the training data in these language pairs is not enough? Collecting data in this aspect will be useful for us to improve the performance of machine translation engines.

**Dynamic Quality Framework**

The abovementioned new benchmarks are dynamic and each one depends on many other factors such as industries, domains, processes used, technologies applied. This also means much more translation quality related data would be generated in each quality evaluation process. Translation evaluation used to be done manually, subjectively and qualitatively. But how can we continue to handle it this way with such a big amount of data? We will have to leverage

---

Many companies started or are interested in collecting data on translation error patterns and productivity scores for a specific translator or vendor.
translation quality tools or software applications to collect, manage and analyze this data.

While each company might be able to develop its own metrics for translation quality, collect translation quality related data, and/or even develop its own translation quality business intelligence software applications, we should consider the strength and benefits of sharing data on translation quality in the industry. Translation quality evaluation is challenging and costly for many companies. And the cost of developing metrics and applications from scratch is even higher. We need to think about the value and power of sharing metrics and data.

At the TAUS Annual Conference and the Roundtable in Washington D.C., TAUS introduced its Quality Dashboard and the Dynamic Quality Framework (DQF) tools. TAUS Quality Dashboard was launched in June 2015. It is an industry-shared platform that helps providers and buyers of translation as well as translators to measure and benchmark quality, productivity and efficiency. Through an open API users and translation technology providers can connect with the Dynamic Quality Framework (DQF) tools and track performance on the Quality Dashboard.

Many companies and translation technology providers have used DQF or incorporated it into their tools. In the panel discussion titled *Treating Translation Quality Metrics as Business Intelligence*, moderated by Clove Lynch and Eduardo D’Antonio (VMWare) at the TAUS Annual Conference in San Jose, the panelists shared their experiences of using some cloud-based translation quality platform such as TAUS Dynamic Quality Framework (DQF). They covered the topics of translation speed, quality for a small number of readers, budget, audience, lifecycle, visibility and content type.

**Quality in Machine Translation**

Translation quality is an important parameter for machine translation. MT suppliers have to measure the quality and control the product of machine translation.
In the panel discussion titled *Managing Different Levels of Quality* moderated by Don DePalma in San Jose, the panelists talked about the buying preferences for human, post-edited, and raw MT from the perspective of quality. From the panel discussion, we can find MT has different criteria in measuring translation quality. For example, we can measure the quality of the MT results in terms of customer conversion. In other words, if more customers in the local market take actions to purchase the product after reading the translation, it means better translation results. The panelists noticed that for many potential customers, accuracy is often more important than fluency. In these cases, people purchase the items even though they don’t appreciate the MT translation quality if they get the main idea of the translation.

TAUS’ Dynamic Quality Framework also provides functions to compare the quality of different MT engines.

**Closing Remarks: Datafication of Quality**

The new translation industry involves more dynamic benchmarks for translation quality and more role players in the translation process (e.g. translators, LSPs, clients, machine translation users, post-editors). We need to take these variables into consideration when judging and controlling translation quality.

There is not a one-size-fits-all translation quality standard in the translation industry and translation quality evaluation results may vary depending on the variables mentioned above. This also means the amount of translation quality related data would grow at an exponential rate. In addition to linguistic data that is used as the reference criteria for translation quality, quality related data per vendor, per language, and per content type would also come into play. Statistics on productivity and technology comparison should be considered in quality evaluation as well. Ultimately, we will need a platform to manage and analyze the data in order to help evaluators make the right decision on translation quality evaluation.