



Weighting Machine Translation Quality with Cost of Correction

2009/07/17 by i18nguy

Machine translation (MT) was one topic yesterday as I enjoyed lunch with Ben Cornelius and Ray Flournoy of Adobe Systems. Of course, a key criteria for evaluating these tools is the quality of the translated output. Most people would say it is the most important criteria. Speed, cost, and integration with other tools are also significant.

However, evaluation of any tool should take the intended application into account. Some applications can use MT output directly. Many more require review and editing of the output before publication. Manual review and editing introduces labor costs and delay which can be significant.

Therefore we should look at the cost of operating the tool plus the cost of post-editing, when evaluating MT tools. Clearly, the optimum is to have 100% quality and no post-edits are required. But, this is not usually the case...

Not all editing tasks are the same. Some edits are easy to make and low cost to fix. Others are labor intensive. The entries that require editing may be obvious and therefore easy to find. Some may be subtle and require more intensive scrutiny to identify.

The post-editing needed by machine translation output will follow a pattern that varies with the MT engine (and its rules, or training, etc.). (Human authors also

• i18nGuy

• Blogroll

- [Common Sense Advisory](#)
- [English](#)
- [False Advertising](#)
- [GlobalSight](#)
- [Got Unicode?](#)
- [i18nGuy blog list](#)
- [i18nGuy Web Site](#)
- [John Battelle's Search blog](#)
- [Lingoport](#)
- [Localization 2.0 \(Lionbridge\)](#)
- [Online Dating Insider](#)
- [Penn State's Computing with Accents, Foreign Symbols & Scripts](#)
- [Unicode Conferences](#)
- [Unicode Consortium](#)
- [W3C Internationalization](#)
- [Web Globalization \(Lionbridge\)](#)
- [Weitzner: Open Internet Policy](#)
- [XenCraft](#)
- [Yunker's Global By Design](#)

have a writing pattern requiring particular classes of edits.) Typographic and terminology substitution errors may be easy to address. Some grammar and style errors may be more costly. Consistency, flow and the relationship among sequences of sentences may be harder yet.

This suggests an interesting criteria for evaluating tools, the joint editing productivity and total operational cost of using the MT tool. An MT product that generates text needing edits that are both easy to find and to fix could be very low total cost. Another tool producing higher quality linguistic output, might still be less productive if post-editing is difficult.

A good metric for MT tools would be to assign a weight proportional to the cost of fixing a problem to each class of error. A document could have 100 typos and be much cheaper to ready for publication than a document with only a few consistency or contextual errors that required thought and consideration to address.

This metric would also help with process configuration. For example, if I have to produce both Mexican and Iberian Spanish translation, based on English source material, I have several options.

If ">-MT->" represents a machine translation step, and ">-PE->" represents a post-edit step:

	Option	Step 1	Step 2
A	Simple MT, then PE	en >-MT-> mx en >-MT-> es	mx >-PE-> mx2 es >-PE-> es2
B	mx to es	en >-MT-> mx mx >-MT-> es	mx >-PE-> mx2 es >-PE-> es2
C	mx post-edit to es	en >-MT-> mx mx >-PE-> mx2	mx2 >-PE-> es es >-PE-> es2
D	es to mx	en >-MT-> es es >-MT-> mx	es >-PE-> es2 mx >-PE-> mx2
E	es post-edit to mx	en >-MT-> es es >-PE-> es2	es2 >-PE-> mx mx >-PE-> mx2

The scenario that is most effective is the one requiring the least editing. This may

not correlate with unweighted measurements of each machine translator's linguistic quality.

When I mentioned this, Ben recalled a demo by ProMT that the three of us attended recently. ProMT machine translation has a nice feature for managing placeholders used to represent program variables.

Here is an example sentence with two placeholders represented by an identifier in curly brackets.

"The file {0} contains {1} words."

The filename and word count would be substituted at run-time for {0} and {1} respectively.

Many machine translation tools segment the text in between the placeholders rather than treating the placeholders as part of the syntax of the sentence. Therefore the placeholders are not properly addressed in the translated output. The problem is exacerbated by tools that convert markup tags to placeholders.

For example, according to **Alex Yaneshevsky of ProMT**, Idiom WorldServer converts:

```
<i>My name</i> <u> is </u> <b>Alex</b>
```

```
{1}My name{2} {3} is {4} {5}Alex{6}
```

The resulting translation gives:

```
{1}{2}Меня зовут Alex{6}{3}{4}{5}
```

Even if you don't read Russian, you can see that "Alex" should retain placeholders as "{5}Alex{6}".

Post-editors must remove the original placeholders from where they are positioned in the text and insert placeholders into the correct locations. This would be a significant cost consideration for either software or markup localization.

ProMT treats the placeholders as part of the sentence resulting in better placement in the output. This simplifies post editing and improves productivity.

(I am not commenting on ProMT translation quality. For this scenario their output

significantly reduces post-editing cost.)

Ideally machine translation would deliver 100% quality. However, if the quality is less than 100%, then evaluating the combination of machine translation and post-editing effort is a more useful measure than selecting tools or configuring workflow based on just quality metrics. Higher quality might be irrelevant if it is more challenging for the human post editor to correct the text.

Tags: [machine translation](#), [quality](#), [metrics](#), [cost](#), [Localization](#), [ROI](#)
Posted in [Localization \(L10n\)](#), [Machine Translation \(MT\)](#), [translation](#) | [4 Comments »](#)

Quality claims for professional vs. crowd sourcing translation

2009/07/01 by i18nguy

To address the debate about professional vs. crowd sourcing translations, I would like to offer a recent experience I had in an unrelated field.

I brought some 50 year old pictures that I wanted scanned to a photography store. I could have done the scans myself, but the store was offering a very inexpensive promotion.

Before I would let the store have the work, I spent 20 minutes with the owner vetting them and establishing my requirements. My number one priority was that the **originals remain undamaged**. Not only did I emphasize this multiple times, but I reviewed the scanner and its feeder, etc. In fact, we made changes to how the pictures were collected after scanning, to assure the pictures would not be bent as they exited.

My second requirement was timing. My deadline was a plane flight a few days later. I also went over resolution, color accuracy, pricing and some other fine points with the owner. He provided all sorts of assurances and testimonials. The scanner was state of the art and very expensive. He would handle the scanning personally to protect the pictures. I questioned how often they did projects like this and he described numerous similar projects and their extensive experience. They were professionals.

As you have no doubt guessed by now, the process broke down, in multiple ways.

1. The scanner glass and reader were dirty causing the images to have lines and marks on them. I noticed this and they redid all the scans the next day.
2. Some pictures were damaged. They weren't bent or mangled, but ink from some of the darker pictures transferred to the scanner rollers which then transferred ink smudges onto pictures of a pure white wedding dress.
3. The scanned images are networked to a server for making CDs and other processing. The server CD burner stopped working. To resolve the problem they transferred the files to another machine to make the CD. This machine had different software on it which for some reason reduced the resolution from 300 dpi to 90 dpi. It also cropped some pictures (for a reason that is not understood) lopping off heads and other undesirable changes.

You can imagine how upset I was at the damaged originals. Later I reviewed the CD and discovered the low resolution and cropped images. I spoke with the owner. The scans were still on the server so they made a new CD.

I'll leave out the remaining details. Suffice it to say that I returned to the store each day 3 more times until I finally had uncropped images of the right resolution, (but showing the now ink-stained wedding dress.) I missed my deadline and spent as many hours with the store and owner as I would have spent if I had scanned the images myself.

The owner, throughout this, was surprised and upset by the problems, immensely apologetic, worked overtime to satisfy me and brought in other staffers to rush through equipment and other fixes and to be as timely as possible. He spoke, I believe sincerely, of honoring his commitment and trying to make me whole, and happy.

Now I hope some of you see the relationship with translation.

Many organizations and individual professional translators in the industry are protesting the use of crowd sourcing. The claim is that quality will suffer.

I know many of the folks in the industry and I do not question the training, skills, attention to quality and passion that goes into providing good translations. As with the store owner, it is a matter of personal pride, that the work be excellent and the customer satisfied.

However, if we **look at the user community and their satisfaction**, we see that the intent does not become achievement. Translation is not just a product of the wordsmithing of an individual. It is a process involving several people, tools and equipment, different kinds of both source material and expected outputs. There are many potential points of failure.

Every experienced translation client has stories of missed deadlines, broken promises, and translations that were rejected by end-users. Many of you will assert that nearly all of the problems with the scanning project could have been managed better and either prevented or anticipated with contingency plans. The same is true for translation projects that go awry. Nevertheless many do go awry.

There are going to be projects that require special skills and attention. These are best attended to by professionals with the relevant experience, and not just professional translators, but organizations that are attentive to the entire process.

However, today the industry frustrates its clients with mediocre project management, inadequate workflow and translation memory tools, poor IT practices and lack of interoperability. It delivers translations that are rejected by user communities with surprising regularity. As long as the failure rate is high enough to cause distrust, clients are going to consider the do-it-yourself solution (i.e. crowd sourcing).

It is a realistic solution as well. For the games market, the burgeoning social networking market, and other markets, end-users are able to self-select the most desirable terms and phrasing. Professional translators do not have a particular advantage here.

For some of the languages of Africa and elsewhere, there aren't sufficient translators, or established glossaries that professional organizations can claim an advantage either.

Clients will not believe that quality will universally suffer under crowd sourcing until the industry improves reliability of its professional services overall and clients are

comfortable that they will get value for the dollar.

I know I won't be using a professional scanning service until I have a need that I can't fulfill on my own.

Tags: [crowd sourcing translation professional](#)
Posted in [Localization \(L10n\)](#), [translation](#) | [Leave a Comment »](#)

About Twitter's mid-day outage in support of Iran bloggers

2009/06/24 by i18nguy

There is additional significance to the mid-day Twitter outage that occurred June 17, 2009 that should not be missed.

Many parts of the world are used to mid-day outages of an America-hosted site that periodically comes down for maintenance.

If 2 A.M. in the U.S. is during your business day, then your business is frequently affected by these outages.

It may only be an hour, and it may be a predicted outage, but it can still cause pain. The inherent message that your region is not important to the owning company is a motivation to consider other providers, where they exist.

Today Twitter was down during the American afternoon, intentionally in preference to supporting the Iranian tweet community protesting the Iranian election results. Let's separate the good intentions for this move and consider the event by itself.

This is a very rare (if not a first) instance of a region being shown preference over the large American market.

On the one hand, equality of languages, cultures and regions is a key principle behind internationalization and is to be welcomed.

It is newsworthy if this attitude becomes a new trend where social networks must consider their worldwide users and give balanced support.

At the same time, if Twitter is also a business application, then its business users require consistency of support and reliable availability.

So as social and business network media become significant across the world, the conflict of which markets suffer through these necessary outages will become more important. This points at an issue of significance behind today's outage.

To be relied upon, Twitter and other social media must adopt techniques that guarantee worldwide uptime. Maintenance methods that provide for continued service as improvements or repairs are made, must be employed.

It cannot be ignored that many American and other Twitter users were anxious not only about Iran, but also the other topics they were discussing and perhaps necessary for business or other reasons, during the hour Twitter was down.

How many more mid-day outages will Twitter users accept before they shift to other more reliable services?

Not many. There is no time of day where large communities of users are not hurt by outages.

Social networks need to get behind 24x7 uptime and phase in maintenance.

Posted in [Uncategorized](#) | [Leave a Comment »](#)

International Domain Name Examples and #idnfail Apps

2009/06/18 by i18nguy

At the [April 2005 Unicode conference IUC27](#) in Berlin there were several interesting sessions on **international domain names (IDN)**.

(The paper I presented at the conference on internationalizing domain names is at www.i18nguy.com/markup/internationalizing-web-addresses.html.)

Here are **example international domain names** that exist today for museums:
<http://www.i18nguy.com/markup/idna-examples.html>

The list is handy for **testing** IDN in applications.

(You may need an appropriate font to read them.)

Why the list isn't in this blog

Originally, the list of **IDNs** were in this blog. It turns out that **WordPress** treats the domain names as URLs and converts the bytes to hex encoded values. This is wrong for IDN. So you **cannot properly link to international domain names in WordPress** blogs at the moment.

As a result, I suggested on **twitter** to use the **hashtag #idnfail** for people to report blog, twitter and other applications that are broken with respect to international domain names.

The initial list of #idnfail apps (which I have not verified myself for each entry) is:

- WordPress
- tweetie
- twitteriffic
- Facebook sharer app
- <http://txtn.us>

You can search twitter for #idnfail to get further updates, although I hope to update the above list from time to time.

Tags: [Berlin](#), [conference](#), [i18n](#), [icom.museum](#), [IDN](#), [IETF](#), [international domain names](#), [internationalization](#), [IUC27](#), [Unicode](#)

Posted in [International Domain Names \(IDN\)](#), [Internationalization \(I18n\)](#) | [3 Comments »](#)

[Blog at WordPress.com.](#)
[Entries \(RSS\)](#) and [Comments \(RSS\)](#).

