



Panel Presentation on Cloud Computing, Crowdsourcing & Machine Translation

LTD General Meeting, ATA Conference, 2010

By Emily Tell

A lively and energetic panel presentation at the general

meeting of the Language Technology Division at the ATA conference addressed the technology trend issues that most bedevil the translation community. Cloud computing, crowdsourcing and machine translation are some of the emerging trends that affect translators in their daily lives, in ways that range from their contracts with

I. Cloud Computing

The first speaker, Beatriz Bonnet, President and CEO of Syntes Language Group, spoke about **cloud computing**, a term used in very different ways, but mostly as a marketing term that denotes processing power in a lot of different servers. This concept, which has been around ever since we have had virtual private servers, is more applicable to companies than individuals in general. However, with the increase in processing power, cloud computing is now a little bit more difficult to define, as it generally takes place on multiple servers. The location of the data, its destination and how it's being backed up is more invisible to people using cloud computing.

In addition to defining **cloud computing** (the processing power in a lot of different servers), Beatriz also defined the terms **service cloud** and **translation cloud** and described how applications such as SaaS (Software-as-a-Service) always reside in the cloud. An example of a translation cloud is what TAUS (Translation Automation Users Society) is doing: a group of companies have come together to form a data repository of translation memories. Such companies as Cisco, Adobe and Microsoft are members of this society, which has created a legal framework for sharing all their translation memories with each other. Paying members of the society have the right to download the translation memories for their own use. The idea is that one company may have the need to translate something already translated by another company, so they avoid "reinventing the wheel." The TAUS Data Association (TDA) shares all this information and has just opened it up to the translator community – not for downloading the translation memories but for consulting and using existing translations. Some issues to look out for are the quality of



Laurie Gerber (left) and Rosana Wolochwianski

clients to the legal and ethical ramifications of using online translation memories. Each speaker gave a five-minute presentation focusing on one example of the phenomenon. Each presentation was followed by a question from the moderator and then a question from the audience. This article will address the first two presentations, since the third presenter, Rosana Wolochwianski, wrote an article about her own presentation. Rosana's article can be found in the next edition of the LTD newsletter under the title, "An Overview of the Impact of Machine Translation on the Professional Translation Community."

the translation memories and the possibility that all the translations might be moved to the cloud, thereby replacing translators. An example of cloud computing outside the translation industry is salesforce.com, which is a product of the Web 2.0 scenario.

How does this trend in what big companies are doing affect the life of freelance translators?

While TAUS is what the big companies are doing, Wordfast and Google Translate are creating big translation memories that freelance translators can contribute to. There is a question as to whether freelance translators are willing to share their translations and contribute them to translation memories. The speaker stressed the importance of being careful about sharing and about reusing translation memories to avoid retranslating, because of the potential for violating contracts between translators and direct clients and/or translation agencies. It is easy for translators not to pay attention to the contracts they have signed. If the contracts are well written, there will probably be a clause that says that everything they do is considered “work-for-hire,” and therefore they don’t own the translations provided. Furthermore, translation contracts normally have clauses stipulating that everything we do is confidential and that the translation is owned

by the translation agency or by the translation agency’s client. If we are thinking about uploading our translation memories to Wordfast or to any repository in the cloud, it is important to ask ourselves beforehand if we really own them, if we have the right to them. Since our

translation memories probably contain a lot of company names, uploading them could be deemed a breach of contract between the translator and the translation agency, and also a breach of contract between the translation agency and its client, which could be a software company which hasn’t released its product yet. Also, translators must ask them-

selves if they are commingling translation memories from a lot of different clients. It may be okay to share the data from one client, but not okay to share the data from the other three clients. The other point is that there are a lot of resources in the cloud that translators can consult—for example, the TAUS translation memories. Other resources such as Translated.net, glossaries and forums are available to translators. Be careful to understand where a translation comes from and don’t trust it automatically. Just because there is a lot of information out there, and it was worked on by other people, sometimes we automatically think it is good, and don’t think it over. If translators find data from the top 10 Fortune 500 companies, don’t assume it’s good. Remember that whatever translators get has to come from the right source.

Relevant links:

www.tausdata.org

www.translationautomation.com

www.translated.net

www.wordfast.net

Google Translate: www.google.com/ig

II. Crowdsourcing

Naomi Bear, the second speaker, is the Director of Microloan Translations and Review at Kiva and she spoke about crowdsourcing and Facebook.

Crowdsourcing is a new term which is defined as the act of taking a task traditionally performed by an employee or a contractor and outsourcing it to a generally unknown large undefined group of people in the form of an open call. This definition is taken from Wikipedia, which is the best example of crowdsourcing. In the case of Wikipedia, a large group of people have come together to iteratively create a body of content. Crowdsourcing is about three years old. It has been around for a long time, however, although in a different form. Sun Microsystems has been doing crowdsourcing for the translation of its open source software for a number of years. Crowdsourcing is a growing trend in the industry, and there are a number of organizations and companies that have begun to experiment with this in the last year or so. In

Before uploading our translation memories to any repository in the cloud, it is important to ask ourselves if we really have the right to them.

the case of translation, the definition could be a little bit broader: An undefined generally large group of people that may vary in terms of size or amount of definition. Facebook is the example which is most talked about these days. It is a very specific type of crowdsourcing and it represents just one point on the range. The way it works at Facebook is that there is an open crowd which is undefined. The only limitation is that you have to be a member of Facebook. Once you're a member of Facebook, you can download the application, contribute translations to their user interface, and participate as you wish. The model for Facebook also works in an iterative way. That is, Translator 1 might contribute a translation into French to befriend somebody. Then Translators 2, 3, 4 and 5 might contribute an alternate version of the same translation. These versions would then be reviewed and voted on by the crowd until one of them is

Other points along the spectrum are the open versus closed crowd and everything in between. There is an open crowd with the unrestrictive Facebook model, but there are other organizations like Asia Online, which is running a project to translate all of Wikipedia into Thai. Asia Online has a system where they initially differentiate their users through some automated testing. Then the users acquire a history, as the paid in-house linguistic reviewers edit the translations to determine which of the contributors of the crowd are more or less reliable.

There are some systems where it's a fully closed crowd. The program being run at Kiva has human testing in which the translators take a traditional translation test. The test is evaluated by humans and either accepted or rejected. If translators are accepted then they have free reign to participate in the translation crowdsourcing program.

The other point on the spectrum that differentiates translation crowdsourcing models today is the review mechanism. There is a perception that no crowdsourced translations are reviewed and that there is 100% reliance on the crowd to identify the best translations, regardless of the crowd's experience and background. In some cases this may work relatively well, as in Facebook, where the elements are relatively small and the crowd is very familiar with that environment and terminology. However, most companies are actually doing some sort of professional review on top of that. So even in a Facebook model, with a voting, ranking system, most companies are paying their linguistic reviewers in top tier languages. So there is an element of professionalism in most translation crowdsourcing today. Organizations and companies that are seeing the value of community involvement still understand the importance of having quality. There is a range and we're still learning which of these models works best.



Naomi Baer (left) and Beatrice Bonnet

approved. The other defining aspect of the Facebook model is that it is translation for very specific segments of the user interface. For example, strings that are short and relatively small in volume – such as 35,000 strings – are being translated into about 65 languages now.

Why are companies using crowdsourcing for translations and taking the work away from professional translators?

According to Naomi, there is a concern and perception that the motivation behind crowdsourcing is simply to cut translation costs.

This isn't true. A recent industry insights report on crowdsourcing stated that one of the key findings was that companies were not implementing crowdsourcing to reduce their costs. There is a real cost to setting up and maintaining an infrastructure like crowdsourcing, and the community needs to be recruited, motivated, sustained. Quality needs to be monitored. The top motivation for companies is community engagement. For an organization or product that has a passionate user base, this is one more way that users can be involved. It is a way for them to participate, stay engaged and define the terminology or the atmosphere that they're working in.

Emily Tell is the CEO of Tellmealgo Translations, Inc. and Assistant Administrator of the ATA Language Technology Division. She can be reached at emily@tellmealgo.com

There is also the issue of market reach. Many crowdsourcing translation programs for software products, for example, are creating translations in a vast number of languages that would not be supported by paid translation. Companies can begin to access new markets in some cases, test the translations, and see if they are viable. Then perhaps they will make a case for professional support.

On the flip side of reach there is the question of access. Going back to the Asia Online project, project participants talk a lot about the fact that internet users in the Asia Pacific region are expected to reach 47% of all internet users by the year 2013. At the same time, only 13.8% of internet content is in those languages, and non-CJK content is only 0.03%. There is real information poverty, and some of these crowdsourcing programs are providing access to information. At the same time that they are enabling people in information-impovertised markets to access all kinds of paid and unpaid content online, they're also making it possible for organizations and companies to reach these markets to sell their products and services.

Another reason why translation crowdsourcing is taking place is time to market. Unlike the traditional sequential localization model, there is a lot faster turnaround. An additional reason is that we can translate more. There is a lot of additional content being

translated that organizations would not budget for. Even organizations that have large translation budgets for their traditional projects and user documentation are using the translation crowdsourcing model. There is one case of a software company that pays for professional translation of all of their products, but also has a large base of user-generated development documentation that they would not translate professionally. Their user base has asked for access to this content in languages other than English and they have begun to support a crowdsourcing model.

So we're not seeing crowdsourcing taking away very much translation of existing content, but are really beginning to see an expansion of the pie in terms of what is being translated and which languages are being translated. In addition, we are also seeing possible new roles for professional translators, as paid reviewers of community translated content.

Relevant Links:

www.kiva.org
www.facebook.com
www.asiaonline.net
www.wikipedia.org

Conclusion

The panel presentation addressed very important issues affecting translators today. Big companies are using cloud computing to create repositories of data located in invisible online servers so that they do not have to invest in new translations and can share existing ones. On an individual level, translators need to be aware of the legal and ethical consequences of sharing translation memories, as there may be contracts in place with their clients regarding the ownership of these translation memories. The phenomenon of crowdsourcing is also a source of concern for professional translators, who may feel that work is being taken away from them in the interest of developing online communities around a brand or website like in the case of Facebook. Translators need not hop on the wagon or sit in silence. Rather, they can observe how these trends are unfolding. With this awareness, they can make educated decisions based on fact, not illusion. □