

# An Overview of the Impact of Machine Translation on the Professional Translation Community

By Rosana Wolochwianski

## When I was asked to contribute to the “Cloud, Crowd,

Machine Translation” panel organized by the Language Technology Division (LTD) at the 50th ATA Conference in New York, I tried to figure out how to offer a snapshot, a brief and concise account that would help us understand where we are standing in relation to MT, and thus be able to consider to what extent it is occurring in isolation, or if it is linked to the other two trends.

we live, the quantity of information that circulates is so huge and is needed so urgently, that in many situations there is no human capacity to respond to that demand, or there is no budget to pay for it, or no time to wait for it. So, for many different reasons, MT has started to play a role in the translation scenario. An alternative to “zero” translation has emerged. In a way, the “let’s make do with less” approach has been adopted. The market has embraced “usable” machine translation, instead of pursuing a “perfect” one.

*In a way, the “let’s make do with less” approach has been adopted. The market has embraced “usable” machine translation, instead of pursuing a “perfect” one.*

In order to get a grasp of this phenomenon, we should understand at least minimally how the different MT systems work. At this point, there are mainly two MT systems available: “rule-based” and “corpus-based” ones. In simple terms, the rule-based systems are the classic ones. Basically composed of a set of

As a translator, whenever I reflect on the technological changes in our industry, most of my reflections are geared towards answering mainly two questions. The first: Does this have a direct impact on our work? The second: What should my stance be, if appropriate, before my peers and the community in general? When it comes to Machine Translation (MT), there surely is no easy answer. I will try to share my views along these two lines.

## Current Status and Use of MT Tools

For over 50 years, researchers have been struggling to create tools that could translate with the same level of quality a human being does. This, obviously, has not been possible, for several reasons, which cannot be covered here. However, in the current information age in which

grammatical rules and dictionaries, they are supposed to analyze language just as humans do, at a lexical, syntactical and semantic level, and provide a target language version of a similar nature.

Corpus-based systems, on the other hand, work with statistics and examples. Since the ‘90s, huge amounts of aligned bilingual material (“corpuses”) have been developed through the use of Translation Memories (TM). The principle behind corpus-based systems is feeding MT systems with these corpuses of bilingual data for training purposes, so the

**Rosana Wolochwianski** is an ATA-certified (E-S), Licenciata in Translation from Rosario, Argentina, where she directs a Center for Translation Studies called CEIT. She is doing research on MT and has presented her findings in several national and international events in the last few years. You can reach her at : [rosana@rwtraducciones.com.ar](mailto:rosana@rwtraducciones.com.ar).

tool can predict, on the basis of a previous translation, how a new translation could be solved. That is the latest tendency and that is why TM repositories are becoming so necessary. Bilingual data has become the most precious fuel for these systems.

In any case, neither of these systems really “translates,” as translation requires the human capacity of inferring and conveying meaning in discourse. However, as the result of rule-application or matching and prediction efforts—or both, in the case of so-called “hybrid” systems—they do render a proposed target language equivalent, which generally requires human review.

In what way does the emergence and popularization of these tools affect translators? We can try to explore the situation by examining the following alternatives: MT with/without Translator Involvement; MT at Company/Freelancing Level; MT as a Free/Paid Resource.

#### **MT without Translator Involvement: as a Free Resource**

There are now several free MT tools available online. They do not affect us as professionals. Those are used by the general public for information retrieval purposes, as an alternative to having NO translation at all. We

should understand that, in general, they are applied to materials that would never have reached the hands of a translator anyway, either because they are needed instantly, or because that public would not pay for a translation service. People use free MT tools to navigate, to make searches, or to get a rough idea of what they are reading in a foreign language for informal or

private purposes, or for internal circulation of perishable documents. The results retrieved are awkward, but as they are free, there seems to be no high expectations, and also no grounds for overly positive or overly negative

claims, legal and otherwise. The public is starting to become aware already (through trial and error) that these raw results are far from perfect. This reality has come to stay. We as translators are not involved, and all we can do is warn the community that these are not reliable translations and should not be used as a basis to make business, ethical or technical decisions.

#### **MT at the Company Level, with Translator Involvement: as a Paid Resource**

On the other hand, there are also commercial MT programs that are used as a productivity tool and that are integrated with TMs and other automation tools in translation companies and departments. This reality affects translators directly, as it incorporates a new task in the usual translation practice, that of post-editing MT output. This methodology is applied to large-volume, usually domain-specific projects with a high degree of repetitive text, and it has proven to be a good way to cut costs and accelerate production, when used effectively. There might be no way back in certain sectors, like the automotive industry, for example. However, we should understand that this methodology cannot be applied to all scenarios, and so it will not replace translators. At this point, even MT tool developers admit that it is not reasonable to think of MT without humans—it is all about making MT + human collaboration work. Besides, managing MT projects requires lots of effort and investment: tools acquisition, time to train the systems, large availability of bilingual corpus and/or glossaries, recruiting and/or training post-editors, etc. Bad implementation could lead to a loss instead of a profit. This model is in its initial expansion phase, experiencing some defeats, but also several success stories.

Where do we stand as translators in the face of this new situation? The tools themselves are not a problem. When a project is well managed, translators and companies reach an agreement that is mutually satisfactory in regards to expected volumes and compensation. As translators, we should understand that this is not taking work away from us, because the volume of translated

*At this point, even MT tool developers admit that it is not reasonable to think of MT without humans—it is all about making MT + human collaboration work*

information is growing exponentially. We can either embrace this new work modality as a new job opportunity, and try to get familiarized with it, or we can choose to work in different fields, in which MT is not applicable.

### **MT at the Autonomous Freelancing Level: as a Paid Resource**

In regards to the use of MT by the freelance community, I think that it is still not happening in a noticeable way, due to many factors. The main one is that translators in general enjoy translating much more than post-editing MT output. Besides, not many translators work in a specific domain at a large enough scale to justify the investment and the time and effort needed to customize the tools. Until recently, only rule-based MT tools have been available for individual translators, as the cost of commercial corpus-based tools is prohibitive for an individual. It would be good if MT tool developers could try to extend their reach into the freelance community by offering affordable and friendly products for those willing to adopt this new technology and be involved in the whole process, and not only in the post-editing phase. The experience of using a product of their own might help translators feel less “estranged” by this technology, in contrast to what happens when they are only presented with the raw output to be post-edited, and have no opportunity to see the whole process. Their feedback might help improve the programs as well, thus boosting collaboration.

### **MT at the Autonomous Freelancing Level: as a Free Resource**

With the recent emergence of the free Translation Toolkit offered by Google, a new participant has entered the scene. And it seems to pose a double-edged sword. This Toolkit is an on-line translation platform in which anyone can upload files for translation, as well as glossaries and/or TMs. In addition to working as a TM translation tool, it also renders MT results for untranslated text. It works as a corpus-based system. That means that all the translations done with it are used to train the system further.

We should all be aware of the fact that we are helping Google train their system with every translation we process, even if we choose not to share the TMs we upload. That’s the logic behind their corpus-based system. As a result, we face a problem of confidentiality. Google might be a resource for semi-professional or non-professional translators. However, as professionals, can we paste the texts our clients provide us for translation on an online tool? I think we have to draw a line here. We usually sign confidentiality agreements. Not long ago, I read a published case of a company that distributed its own customized MT tool to its overseas employees who were not English speakers, so that they could read the company’s bulletins and intranet site and would not have to “cut and paste” sensitive information into public MT tools. This issue of confidentiality lies at the heart of our profession, and I think we need to adopt a firm stance in this regard.

### **MT’s Place within the “Cloud, Crow, MT” Trilogy**

MT is certainly not isolated from other trends in the industry. With bilingual corpuses having become the “fuel” for many current MT systems, TM sharing is now part of the MT agenda. Companies can use their own TMs or get them somewhere else. There are several ventures related to TM sharing. Probably the most notable is the TAUS Data Association (TDA), an attempt to pool and share data from the largest MT users in the industry in order to feed their MT systems.

On the other hand, by offering translation tools for free (at least for a while, as a disclaimer warns that it might be charged in the future), Google might amass one of the largest bilingual corpuses. Interestingly enough, one of the features Google Translator Toolkit offers is the possibility of translating articles that are to be uploaded to Wikipedia or Knol, a task that is usually tackled by crowd-sourcers (a free tool for people who translate for free, leveraging freely shared resources). So, it is easy to see that we are submerged in a highly complex and interwoven professional reality, no less dynamic or interesting, and with a lot more to come and debate. □