

# Collaborative Translation: technology, crowdsourcing, and the translator perspective

Alain Désilets  
Workshop Chair  
National Research Council of Canada  
[alain.desilets@nrc-cnrc.gc.ca](mailto:alain.desilets@nrc-cnrc.gc.ca)

## Keywords

Collaborative Translation, Translation Crowdsourcing

## 1. BACKGROUND

This one-day workshop was held at AMTA 2010, in Denver Colorado, on October 31<sup>st</sup>, 2010. It focused on the impact (present or future) of new collaborative and crowdsourcing technologies on the world of translation.

Collaborative and social networking technologies like Wikipedia, Facebook and Amazon Mechanical Turk are having profound effects in many spheres of human activity. Translation is no exception, as evidenced by the recent emergence of collaborative technologies and paradigms such as:

**Translation teamware:** systems that allow multidisciplinary teams of professionals (translators, terminologists, domain experts, revisers, managers) to collaborate on large translation projects, using an agile, grassroots process instead of the more assembly-line, top-down approach found in most translation workflow systems.

**Collaborative terminology resources:** Wikipedia-like platforms for the creation and maintenance of large terminology resources by a crowd of translators, terminologists, domain experts, and even general members of the public.

**Translation Memory sharing:** platforms for large scale pooling and sharing of multilingual parallel corpora between organizations and individuals.

**Online marketplaces for translators:** eBay-like systems for connecting customers and translators directly, with minimal intervention by a middle man.

**Translation crowdsourcing:** Mechanical Turk style systems for splitting translation projects into small chunks, and distributing them across large crowds of mostly amateur translators.

**Post-editing by the crowd:** systems allowing a large crowd of mostly amateurs to correct the output of machine translations systems, to suggest better translations.

The workshop brought together a multidisciplinary group of researchers and practitioners from both fields of technology and translation, in order to discuss and explore the impact, present and future, of this type of technology. In order to foster constructive two-way dialogue between developers and potential users of these technologies, we used a highly participatory, discussion oriented format, as opposed to a more lecture-style approach.

## 2. WORKSHOP COMMITTEE

The workshop committee consisted of the following people:

Alain Désilets (chair), Research Officer, Institute for Information Technology, National Research Council of Canada

Naomi Baer, Director, Microloan Translation and Review, Kiva.org

Renato Beninato, CEO, Milengo

Chris Callison-Burch, Professor, Center for Language and Speech Processing, Johns Hopkins University

Kyo Kageura, Professor, Library and Information Science Department, University of Tokyo

Elina Lagoudaki, Professor, Humanities Department, Imperial College London

Dorothee Racette, ATA President-elect (2009-2011)

Philip Resnik, Professor, Department of Linguistics and Institute for Advanced Computer Studies, University of Maryland

Willem Stoeller, Director accounts, Lingotek

## 3. PROCEEDINGS CONTENT

In these AMTA proceedings, you will find the content of the two position papers which were submitted ahead of the event, as well as our two invited talks.

### Invited talks

Robert Munro, "*Crowdsourced translation for emergency response in Haiti: the global collaboration of local knowledge*"

Jost Zetzsche, "*Crowdsourcing and the professional translator*"

### Position papers

Kronrod, Y., Resnik, P., Buzek, O., Hu, C., Quinn, A., Bederson, B. B., "*Improving Translation via Targeted Paraphrasing*"

Kumaran, A., Datha, N., Ashok, B., Saravanan, K., Ande, A., Sharma, A., Vedantham, S., Natampally, V., Dendi, V., Maurice, S. "*WikiBABEL: A System for Multilingual Wikipedia Content*"

Because of the workshop's participatory format, the bulk of what transpired at the event could not be published beforehand in these AMTA proceedings, but you can find it documented on the workshop's official wiki site:

<http://www.wiki-translation.com/AMTA+2010+Workshop>

Below is a list of themes which had been submitted a-priori to participants in the Call for Participation. Note however that additional themes may have been raised during the event, and documented on the wiki site.

### **3.1 Impact on the translation profession**

Some of these technologies (translation teamware, collaborative terminology resources, Translation Memory sharing, online marketplaces) present clear benefits and new opportunities for professional translators. But others (translation crowdsourcing, post-editing by the crowd) could present a threat to their livelihood. How can professional translators prepare for these developments? Will these technologies decrease demand for professionals, or will they increase the pie and be used only for content which currently is not being translated at all (for example, by allowing speakers of small languages like Haitian Creole, to translate content that is particularly relevant to them)? Will professional translators still play a remunerated role, even in cases where tasks are crowdsourced to amateurs (for example, by performing quality assurance or coaching the amateurs)? Which skills/computer resources/qualifications are needed for staying in the business of translation in this new context?

### **3.2 Impact on translation technology**

How could these technologies be used to improve the performance of machine translation systems? Can millions of people, professional translators and amateurs, teach machines how to do a better job at translating? Can machines be used to facilitate collaboration between humans, for example by connecting customers with translators who seem particularly suited for a given translation project? In a crowdsourcing context, how should we adapt tools originally developed for professionals, so that they are better suited to the specific needs and limitations of amateurs?

### **3.3 Quality assurance and appropriateness of the technology**

All of the above technologies lead to environments which are more grassroots, and less tightly controlled from the top than is typical found in most professional contexts. This is true even of technologies that specifically target professionals. What effect does that have on quality? How can we characterize circumstances where such collaboration will increase quality, instead of decreasing it? How should these technologies be used in contexts with different quality requirements, ranging from "fit for gisting" to "fit for dissemination" quality? Can quality assurance itself be

done collaboratively? How can tools be designed to make the crowd collectively smarter than its individuals (wisdom of crowds effect), instead of having it act as a mindless mob?

### **3.4 Co-development and mutual understanding between stakeholders**

How do we foster constructive dialog between stakeholders, to ensure that these technologies reach a balance point that meets their respective needs? How can developers learn more about professional translators and their work, in order to build collaborative environments that leverage the unique skills of that constituency? How can professional translators and their customers learn more about the possibilities offered by these new technologies, so that they can use them to improve productivity while still ensuring fair compensation and quality? How can professional translators reach out to translation buyers to make them understand the benefits and limitations of such technologies (e.g., why would it not be a good idea to crowdsource translation of a patent)?