Translators'knowledge in the Cloud: The New Translation Technologies

Johanna Monti

Department of Social, Political and Communication Sciences University of Salerno, Fisciano, Italy, jmonti@unisa.it

Abstract: After machine translation, the translation technology market is witnessing a new revolution. Relevant changes are taking place under emerging phenomena of the Web such as crowdsourcing, i.e., the exploitation of a community/group of people to perform tasks normally performed by employees and cloud computing technologies, which enable ubiquitous access to digital content and online multilingual translation tools.

In particular, the combination of crowdsourcing and cloud models of automatic/assisted translation is taking place on a large scale, combined with the availability of tools shared in translation environments.

This contribution will analyze the impact of the new collaborative translation technologies on the translation process and the working practices of translators, highlighting the possible implications in the field of translation teaching.

Keywords: *Crowdsourcing, Machine translation, Collaborative translation, Translation teaching*

1. The new online translation technologies: interaction of crowdsourcing and cloud computing

In the last fifteen years we have been witnessing a complete turn in the availability of linguistic resources and free machine and assisted translation tools on the Internet. Emerging web phenomena such as crowdsourcing, i.e., the exploitation of a community/group of people to perform tasks normally performed by employees (Howe, 2006) and cloud computing, which allows users ubiquitous access to services and online tools for translation and multilingual digital content, are further changing the scenario. In particular, the combination of crowdsourcing and cloud models of automatic/assisted translation is taking place on a large scale inside collaborative translation platforms.

In the translation field crowdsourcing refers to the use of professional translators and non-professionals to perform typical translation and localization tasks either on payment or on a voluntary basis. Common Sense Advisory, an American market research company, has coined the acronym CT3, or "community, crowdsourcing, and collaborative translation," which collects the different denominations used to highlight the main feature of this emerging phenomenon, i.e., the collaborative aspect within a community of professionals

or occasional translators who belong to a "crowd" of volunteers willing to contribute to translation tasks.[1]

Generally, this practice of exploitation of collective intelligence in the field of translation is performed as follows:

- the documents to be translated are shared on the web. This sharing can occur either within dedicated environments, and it is therefore addressed to a group of professional translators, or on sites open to the public where the work takes place on a voluntary basis and, in this case, it is aimed at non-professional and occasional translators;
- the work performed by professional, occasional and non-professional translators is then submitted to a review process, which can again be assigned to professionals and non-professionals, depending on the type of texts and the purpose of translation;
- professional translators are usually paid in a conventional way, but volunteer translators, working for free, are paid through non-conventional forms of social gratification, such as the attribution of a score in the list of the people who contribute to the translation, up to public recognition of leadership, when they reach the top of the list, or simply the opportunity to acquire new knowledge or learn something new.

The idea of using crowdsourcing in translation is originated by the need to carry out translation projects in a short time. It allows to produce large volumes of translations in a short time, at low costs and of an acceptable quality. Therefore, it seems to be an adequate alternative in terms of costs and quality both to machine translation, which produces large amounts of translations, but of low quality, and to professional translators, who produce quality translations but at high costs. On the contrary, it very often requires the combination of both these elements, i.e., professional or occasional translators edit machine translation outputs.

Since 2006, this form of exploitation of collective intelligence in the field of translation has paved the way to collaborative practices of translation on a large scale, which, on one hand, are based on the active involvement of translators, including non-professionals, usually to localize open-source products and online platforms and, on the other hand, on the voluntary feedback by users about the quality of machine translations.

Examples of this alliance are now widespread, but the true pioneers of this practice were social networks like Facebook, LinkedIn and Twitter, which were localized in many different languages thanks to the work of their followers. In particular, in 2008, Facebook launched its application *Translations*, in order to localize the interface into new languages and to translate the continuous updates to the platform. In this way, Facebook has been localized in over 70 different languages (with about 100,000 words for each version) at a surprising speed (for instance the entire French version was translated by 4,000 users in 14 hours). The localization and translation strategies used by Facebook are based, on one hand, on the free work of its fans and, on the other hand, more recently, on Microsoft Bing Translate for the translation of posts.

---

[1] http://www.dqglossary.com/simple/thoughtData/3734.html

In *Translations - Go vote on translations* users can choose the best translation among the possible solutions suggested by the system or, if they don't like them, translate from scratch. The social dimension of the activity is fed by the *Facebook Translations Team* group, which is used by the members of the management team to communicate with translators on various technical aspects and in which translators can discuss their problems, ask for tips and give advice on possible translation solutions.

In the abovementioned examples, crowdsourcing is used not only in order to reduce costs, but also to translate in commercially unattractive languages and finally as a means to increase and loyalize users giving them the possibility to shape the image of Facebook according to their tastes and expectations. Thanks to the active involvement of users in the localization of the French, German and Spanish versions, Facebook, for example, recorded an increase that went from 52 to 124 million hits (Britton and McGonegal, 2007; Eskelsen et al., 2008).

Therefore, localization is the main engine of crowdsourcing, since this new way of translating offers considerable advantages for large companies in respect to the localization of website contents and their products, but also to the development of language resources for translation projects and the training of translation software applications. For instance, IBM launched the project *no.Fluent* to build a multilingual parallel corpus using its voluntary employees around the world. One year after the start of the project about 3,000 volunteers had contributed approximately with 36 million words (mainly chat messages and translations done collaboratively), editing the translations made by the IBM MT system.

But localization is no longer the only aim of crowdsourcing, since it is also used in subtitling, e.g. in dotSUB[2] or TED,[3] and even for literary translation, e.g. for the translation of the Harry Potter saga into German.[4]

Crowdsourcing is thus adopted as a novel approach in performing all the different phases of a complete localization/translation process, as highlighted by Désilets (2011), who identifies several forms of crowdsourcing that affect translation, from organizational TeamWares and specialized sites for translation to the availability of platforms for:

- creating and sharing terminology resources and translation memories, i.e., Wiki platforms such as the Worldwide Lexicon Project,[5] an open source collaboration platform based on a huge database of translations usable for any website or web application. Other examples are UrbanDictionary,[6] TermWiki,[7] WeBiText,[8], TAUSData Association;[9]

---

[2] http://dotsub.com/

[3] http://www.ted.com/translate/languages/it

[4] www.had-community.de/HaD

[5] http://www.worldwidelexicon.org/home

[6] http://www.urbandictionary.com/

[7] http://it.termwiki.com/

[8] http://www.webitext.com

[9] http://www.tausdata.org

- distributing parts of large translation projects to professional or occasional translators such as My Genco[10] (for professional translators) or Mechanical Turk[11] (for non-professionals), virtual platforms where buyers can communicate and conduct transactions with translation suppliers;
- providing translations or editing machine translations, such as the collaborative translation environments Google Translator Toolkit, or Geoworkz[12] by Lionbridge.

This latter type of platform, based on the interaction of crowdsourcing and cloud models of assisted translation systems, requires a closer investigation for its impact on translation work practices, the interaction with the translation team and, finally, the working conditions of professional translators.

As an example, Google Translator Toolkit is a free collaborative translation environment, where users can submit their documents to machine translation or machine assisted processes, revise, edit and store translations in translation memories and invite other people (via email) to share the translation or editing work.

Translation memories created by users contain invaluable information for the Google MT engine, which is based on the use of parallel corpora, i.e., original texts aligned with the corresponding translations, stored by users on the platform made available by Google.

There are therefore clear benefits both for users, who can use a free repository where to process their translation work, using what has been previously translated by themselves or by other users, and for Google, which draws on the translations stored in translation memories to improve the performance of its system. However, this is a collaborative environment for occasional translators, as there are limits to the amount of data and formats that can be used, there are no typical translation memories features such as fuzzy matching, and no quality control procedures, and finally it shows data confidentiality problems.

Nevertheless, it was one of the first translation platform of this kind, and it has inspired professional translation collaborative environments that enable ubiquitous access to digital content and online multilingual translation tools within a team, such as Geoworkz by Lionbridge, a fee-paying environment for translation service providers and professional translation companies, based on SaaS-based solutions,[13] which provides access and real-time updates to translation memories, glossaries and features for data sharing within a team, and also between customers and suppliers. More and more software vendors of translation tools are choosing the provision of their products into collaborative translation platforms, among others MemoQ Server[14] of Kilgray Translation Technologies, The

---

[10] http://mygengo.com/

[11] https://www.mturk.com/mturk/welcome

[12] http://www.geoworkz.com/support/training.aspx#Translator

[13] *Software as a service* (*SaaS*) is a software distribution model in which a software company develops, and manages a web application available to customers on the internet, allowing ubiquitous access to products as a fee-paying service.

[14] http://kilgray.com/products/memoq-server

Translation Network by LingoTek,[15] Crowdin,[16] Wordbee Translator,[17] Wordfast Anywhere,[18] XTM Cloud.[19]

2. The impact of crowdsoucing and cloud computing on the translation process and the working practices of translators

The use of these technologies is changing the way in which translation services are offered with a great impact on the process, performed inside spaces which allow fluidity and flexibility and in which translators become proactive subjects, able to use the new translation technologies, and to adapt themselves to new models of collaborative translation.

Recently, many translation scholars have addressed this changed technological scenario, such as García (2009, 2010a and 2010b) and Pym (2011 and 2012). Furthermore, a specific workshop devoted to this issue, *Collaborative Translation: technology, crowdsourcing, and the translator perspective*, was held in connection with the AMTA 2010 conference, and in 2011 *Translation as a Social Activity - Community Translation 2.0,* a volume of *Linguistica Antverpiensia,* was published.

The process of translation is deeply changed by the use of this new generation of translation technologies and in particular by collaborative environments in which the interaction man/machine is particularly significant. Cloud applications offer useful tools to translators, such as automatic/assisted translation tools, glossaries, translation memories, editing features together with software applications for the cooperation among the different actors of a translation process (translators, editors, terminologists, customers and so on), such as Instant Messaging applications.

The first change concerns the use of automatic translation and translation memories in the translation process. Both do not represent an option to translators any more, on the contrary they are now integral part of the workflow. The combination of translation memories with automatic translation and the terminological resources, prepared in the preliminary phases of the translation process, is a main feature in all the various models of collaborative environments available on the web, from Google Translator Toolkit to commercial environments such as Geoworkz or MemoQ server. This means that now the translation process has necessarily to be carried out using the translation technologies: an unthinkable situation until few years ago. As a matter of fact, translation memories were used only in advanced technological sectors of the translation market such as in localization processes (Monti, 2007), whereas machine translation was experimented only for technical translations by large companies or bodies. Nowadays these technologies, integrated into collaborative environments, are used for every type of translation by a large audience of

---

[15] http://www.lingotek.com/

[16] http://crowdin.net/

[17] http://www.wordbee.com/

[18] http://www.wordfast.net/?whichpage=anywhere

[19] http://www.xtm-intl.com/xtmcloud

specialists (translation service providers, professional translators, editors) who have to adapt their interaction and work practices to the new work modalities.

Concerning this point, Kelly et al. (2011) highlight the change from the TEP (Traslate, Edit and Publish) linear model of the translation process to a new model based on the abovementioned translation technologies and cloud computing applications, in which the work is performed at the same time by the different members of a translation team, even on a same document, as it happens for instance in Google Translator Toolkit, where the modifications are made available to all people who share a document. This new way of working is called "parallel translating", which not only refers to the traditional distribution of a large amount of translation work in a translation group, but also to translating and editing the same documents simultaneously and in real time. It shortens considerably the translation process, entailing further advantages such as the availability in real time of the editor's changes or quality controls to translators.

Hence the traditional concept of translation group, based on a vertical management of translation jobs, and in particular of big translation jobs where a project manager organizes the translation process according to the TAP model and where the translators' task is limited to the part of work they have been assigned to, is replaced by the concept of translation community. In the community, translators interact continuously and in real time among peers, giving their contribution in the exchange of ideas, suggestion of best practices, search of relevant information, solution of translation problems. The concept of "community", which highlights the social dimension of the interaction for the achievement of a common goal, was initially used to refer to the communities of occasional translators, who voluntarily joined a translation project. Now it also refers to the communities of professional translators, who take advantage of being members of such communities in different ways: in finding information, in developing language resources (glossaries, terminological resources, translation memories) on a collaborative basis, interacting with the other members of the community.

The community is based on the use of the new translation technologies, so that translators become post-editors of translation produced by machine or machine assisted translation systems. Post-editing becomes, indeed, the main activity of translators, whose creativity, usually used in solving translation problems, is now expressed in quite a different way from the past since it has to take into account ready-made solutions identified by the translation systems. Many scholars have recently analysed this issue from a theoretical point of view (Austermühl 2001, 2006; Corpas Pastor & Varela Salinas 2003; Esselink 2000; Pym, 2003; Torres del Rey 2005), but also with reference to translators'training (O'Brien 2002).

The new element in this context is represented by the fact that thanks to the crowdsourcing used by companies, the translations edited by translators are used to improve the outputs of translation technologies as it was never possible before. As an example, in Google Translator Toolkit, the edited translations represent valuable resources to train the statistical engine of Google Translate so that its outputs become more and more reliable. It is a virtuous circle put in place by translation software developers as highlighted in the following figure:
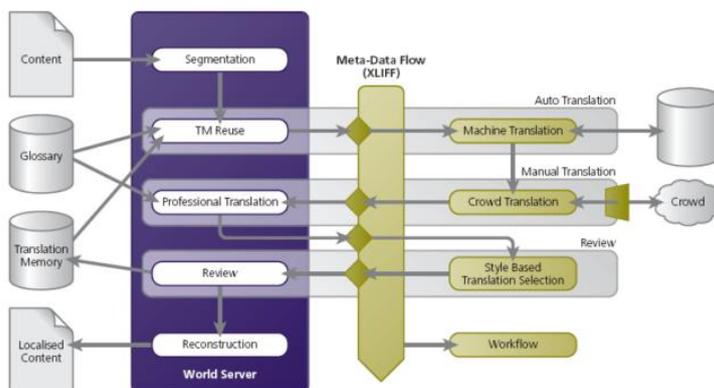
Figure 1 – Use of crowdsourcing in a translation process (Carson-Berndsen et al., 2010)

Translators use machine translation and machine assisted translation to speed up the various stages of a translation task. In this manner they provide valuable linguistic resources which allow to tune the products and can be reused in new translation projects. Translation vendors are increasingly choosing to use and in some cases to develop proprietary collaborative translation environments, in order to ease the management work of translation orders, data control and the quality of the final product. This new scenario very often becomes the only possible way of cooperation between translation companies and freelance translators and this leads to further remarks concerning translators' training.

3. Implications for methods and practices in translation teaching

Translation technologies are gaining a relevant place in academic curricula even if their teaching is not an autonomous discipline, as solicited by many scholars (Alcina, 2008; Fernández, 2003; Piqué, 2002; etc.).

Furthermore, despite the considerable improvements of the new collaborative translation technologies and the extensive changes they have introduced in translation practices, professional translators seems not to be so aware of them, as highlighted by a recent study by Gough (2011). In particular, the most surprising result concerns the translators in their early training stage, who are not well informed about the technological developments in this particular field and whose main source of information is the Internet rather than translation courses. This fact leads us to reflect on how these technological changes, which affect the work practices of professional translators so deeply, have to be considered more carefully in translation teaching and should play a more relevant role in academic translation courses.

O'Hagan (2008), for example, suggests the use of collaborative environments as new possible translation training environments for translators and recently a few teaching experiments in this direction have been already carried out.

An example is the Aula.int project at the university of Granada (Lobo et al., 2007), drawn on the concept of "Professional Approach to Translator Training" (PATT), which develops a dynamic and virtual model of translation based on IT technologies, combining elements

of role-play exercises, tasks assigned to the groups of learners, case studies and simulations.

Renée Desjardins (2011) exploits the inclination of young generations to join social networks and uses Facebook during the translation classes on the assumption that the cooperation among peers and their interaction in social networks are important to prepare the students to work in a virtual environment, helping them to navigate the social, academic and professional spheres, physical and virtual, in which they interact.

Sierra et al. (2011) suggest to adopt a new teaching model, TWITT (Training Web Interaction and Translation Technologies), based on the use of the various translation technologies by translation students to cooperate in a work group, sharing materials, information and knowledge.

The recent free availability on the web of cooperative translation environments in which translation technologies and human interaction are combined shows the way forward to new possibilities in translation teaching, especially towards the simulation of real translation processes.

In 2010, I carried out an experiment in a translation class held at the University of Salerno for the last year of the master degree in Foreign language and literatures, using Google Translator Toolkit. The main aim was to ease the learning process by means of the achievement of translation projects using translation technologies in a working environment, insofar as possible similar to real translation work conditions. I chose Google Translator Toolkit because it is free and it allows to access and use a collaborative environment based on human aided translation tools in a simple and immediate way.

Students were able to understand and put in practice the concept of collaborative translation, in particular by means of:
- complete translation processes (analysis, translation, editing, quality control) with well-defined roles (translator, editor, terminologist) inside a translation community;
- the simulation of real remote work contexts, which is a widespread practice in the professional world, and the cooperation in a translation team by means of instant messaging applications or comments on documents shared inside the group;
- the interaction of machine and human assisted translation tools;
- the development, use and maintenance of a shared set of linguistic resources (mainly glossaries and translation memories);
- the evaluation of the reliability and quality of tools, resources and the final products of the translation process;
- new translation practices such as parallel translating and post-editing of the translations produced by machine and human aided translation.

Students evaluated the possibility to translate using new technologies very positively, since they were able to work as a real translation team, sharing knowledge and doubts in a technologically advanced environment. In this way, the students have acquired knowledge about:
- the translation process, its different stages, the various roles of the members of a translation team, the best practices in the process management,

- the new translation technologies, that up until recently were only accessible paying expensive licenses, as in the case of translation memories or which were not very usable because of poor quality of results as in the case of the free online machine translation services.

4. Conclusion and future prospects

The translation market is changing very rapidly due to the emerging translation technologies based on the use of crowdsourcing and cloud computing. This implies reconfiguring the translation process and translators' work practices as well, since not only new tools and applications are now avalible online, but also new skills and attitudes are required to translators: the translation process is in fact more and more based on collaboration and information sharing. Translation teaching should take into account these emerging trends and should adopt teaching cooperative models based on the new translation technologies, which allow the simulation of real work contexts.

References

Alcina, A., (2008). Translation technologies. Scope, tools and resources. *Target, 20* (1), 2008, 79-102.

Austermühl, F. (2001). *Electronic Tools for Translators*. Manchester: St. Jerome.

Austermühl, F. (2006). Training Translators to localize. A. Pym et al. (eds.). *Translation Technology and its Teaching*. Tarragona: Servei de Publicacions.

Britton, D.B. & McGonegal, S. (2007). *The Digital Economy Fact Book*, Ninth Edition, Washington, D.C.: The Progress & Freedom Foundation.

Carson-Berndsen, J., Somers, H., Vogel, C. & Way, A. (2010). Integrated language technology as a part of next generation localization. *Localization Focus, 8* (1), 53-66. Retrieved 1st February, 2012 from http://www.localization.ie/resources/locfocus/LF_Vol%208%20Issue%201.pdf

Corpas Pastor, G. & Varela Salinas, M.J. (eds.) (2003). *Entornos informáticos de la traducción profesional: las memorias de traducción*. Granada: Atrio.

Désilets, A. (2011). Wanted: Best Practices for Collaborative Translation, TAUS Report. Retrieved 1st February, 2012 from http://www.translationautomation.com/best-practices/wanted-best-practices-in-collaborative-translation.html

Desjardins, R. (2011). Facebook me! Arguing in favour of using social networking as a tool for translator training. *Linguistica Antverpiensia*, *10*, 175-194.

Eskelsen, G., Marcus, A. & Ferree, W.K. (2008). *The Digital Economy Fact Book*, Tenth edition, Washington, D.C.: The Progress & Freedom Foundation.

Esselink, B. (2000). *A Practical Guide to Localization*, Amsterdam & Philadelphia: John Benjamins.

Fernández, M. (2003). *E-Tradumática. Fonaments pedagogics i didactics per a l'enseyament de la tradumática a distancia*. Barcelona: Universidad Autónoma de Barcelona.

García, I. & Stevenson, V. (2009). Google Translator Toolkit: Free web-based translation memory for the masses. *Multilingual*,16-19.

García, I. (2010a). Is Machine Translation Ready Yet?. *Target, 22* (1), 7-21.

García, I. (2010b). The proper place of professionals (and non-professionals and machines) in web translation. *Revista Tradumàtica*, 8.

Gough, J. (2011). An empirical study of professional translators'awareness, use and attitude towards Web 2.0 technologies. *Linguistica Antverpiensia*, *10*, 195-228.

Howe, J. (2006). The rise of crowdsourcing. *Wired*, *14*(06).

Kelly, N., Ray, R. & DePalma D.A. (2011). From crawling to sprinting: Community Translation goes mainstream. *Linguistica Antverpiensia*, *10*, 75-96.

Monti, J. (2007). Localizzazione: il ruolo e i saperi della traduzione. In G. Marchesini & C. Montella, *I saperi del tradurre. Analogie, affinità e confronti* (pp. 173 – 197). Milano: Franco Angeli.

O'Brien, S. (2002). Teaching post-editing: A proposal for course content. *6th EAMT Workshop Teaching Machine Translation*, Manchester, 99-106. Retrieved 1st February, 2012 from http://www.mt-archive.info/EAMT-2002-OBrien.pdf

O'Hagan, M. (2008). Fan translation Networks. An Accidental Training Environment? In J. Kearns (ed.) *Translator And Interpreter Training: Issues, Methods, Debates (*pp. 158-181)*. London: Continuum.

Olvera Lobo, M. D., Robinson, B., Castro Prieto, R.M., Quero Gervilla, E., Muñoz Martín, R., Muñoz Raya, E., Murillo Melero, M., Senso Ruiz, J. A., Vargas Quesada, B. & Díez Lerma, J. L. (2007). A professional approach to translator training (PATT). *Meta, 52* (3), 518-528.

Piqué i Huerta, R. (2002). Apuntes para un modelo pedagógico de enseñanza a distancia de la tradumática, Barcelona: Universidad Autónoma de Barcelona. Retrieved 1st February, 2012 from http://www.fti.uab.es/tradumatica/papers/articles/80_esp.pdf

Pym, A. (2003). Redefining Translation Competence in an Electronic Age. In Defense of a Minimalist Approach. *Meta*, *48* (4), 481-497.

Pym, A. (2011). What technology does to translation, Working version 3. June 2011. Retrieved 1st February, 2012 from http://usuaris.tinet.cat/apym/on-line/translation/2010_technology.pdf

Pym, A. (2012). Translation skill sets in a machine translation age. Paper presented to the 16th International Symposium on Interpretation and Translation Teaching. Graduate Institute of Cross Cultural Studies, Fu Jen Catholic University, Taiwan, January 14-15, 2012. Retrieved 1st February, 2012 from http://usuaris.tinet.cat/apym/on-line/training/2012_competence_pym.pdf

Torres del Rey, J. (2005). La interfaz de la traducción. Formación de traducctores y nuevas tecnologías. Granada: Comares.