



The proper place of professionals (and non-professionals and machines) in web translation

Ignacio Garcia
University of Western Sydney

Abstract

The amount of work available to the professional translator on the web appears to be shrinking, due to the pincer effect of both machine translation and “crowdsourcing”. To maintain its professional status, the translator must choose between specialising in a highly regulated area and becoming a language quality assurance expert.

Keywords

machine translation, crowdsourcing, professional translation, web-aided translation, translation memory.

1. Introduction

The amount of work available to translators on the Internet appears to be shrinking. The web is no longer the publisher-centred hypertext platform of the nineties. It has become a sophisticated tool that shapes the work and leisure patterns of its users. Of recent web developments, often conflated under the Web 2.0 tag, a major one is linked to crowdsourcing, roughly meaning the delegation to (unpaid) volunteers of tasks previously reserved for professionals. Journalism and photography are given as typical examples of areas in which the crowd is replacing the professional. Translation is another. As the amount of professional work available on the web dwindles, translators are not only witnessing an attack from the crowdsourcing flank. They have another flank to contend with: machine translation.

Following translators as they navigate the web between the Scylla of crowdsourcing and the Charybdis of machine translation (MT), this article will place its focus first on assessing the damages to the profession so far and then on scanning the horizon in search of an elusive way out.

2. The calm before the storm

There was an ideal time when the web seemed to offer professional translators ever-increasing amounts of highly skilled work. The only condition was their willingness to train themselves to use translation memory (TM), and to embed themselves in localisation teams, involving fellow translators, computer engineers and project managers.

Crowdsourcing was as yet unheard of. Professionals were blinded by the ubiquity of Microsoft and proprietary software, incapable of noticing the underground movement of Free and Open Source Software (FOSS), which remained faithful to the ideas and practices of the eighties, which indeed involved software development – and software translation – by volunteers. The trend in the nineties was of corporations relying on professionals for software, web development, and translation. Beyond closely related languages (e.g. Catalan-Spanish) or restricted environments (e.g. Canadian METEO), MT could

only make sense for in-bound, gisting purposes – perhaps to find if a particular text indeed deserved proper translation - and translators warmly welcomed it for this purpose.

Early in the decade, reality began to set in. Translating for localisation was not as glamorous as it seemed. It mostly involved the donkey work of managing segments: approving 'exact matches', adjusting 'fuzzy matches' and translating 'no matches' from scratch. This was not even translation proper: issues of context and function, not to mention authorship, so dear to translation theorists, lost relevance against the 1:1 equivalence model of the translation unit sitting in a memory database.

In the early nineties, TM technology was as much the initiative of the computer-savvy freelance translator (e.g. Jochen Hummel and Iko Knyphause for Trados, Emilio Benito for Déjà Vu) as that of the language service provider (e.g. Transit for Star) or the corporation (e.g. IBM Translation Manager for IBM). However, the benefits of using translation memory soon flowed in the corporations direction, as 'Trados discounts', the practice of not paying the full price for exact and fuzzy matches, became commonplace late in the nineties. Amid the tight deadlines, translation and translation related management and engineering activities, first conducted in-house, started to get outsourced, then offshored, pushing the price-per-word down in the process. If translators wished to seek comfort in numbers, they could be sure that offshoring hurt engineers and managers even more.

This slow deterioration of working conditions, which became noticeable early in the decade, gained speed towards the end. The first signal involved translation management systems, now implemented in software as a service (SaaS) mode; then memory and terminology databases, accessed by freelance translators also through the web browser via user name and password. This alone disempowered translators in at least two ways: by forcing them to always work online and with the tool of the client's choice, 'Trados-compatibility' in this mode losing its functionality, and by retarding the speed of database leveraging - by microseconds, but irritating nonetheless. The real accelerators were to be, however, as indicated, machine translation and crowdsourcing.

3. The Scylla – machine translation

In October 2007, Google Translate disconnected the SYSTRAN rule-based engine it had relied upon, and switched on the statistical machine translation (SMT) engine it had built in-house. SMT had suddenly moved from the lab to the most prominent of all spaces: the interface of world's most frequently used search engine. There it went from strength to strength, covering some fifty languages to date. Yahoo's Babel Fish (still powered by SYSTRAN) translates between fourteen only. To spice things up, Microsoft's Bing Translator entered the fray, with its own 'syntactically-based' SMT. Curious translators can check how these three freely available, full web-page strength engines compare for their own language pair and chosen type of text at gabble-on.com, a site which, under its rough looks, hides a new and sophisticated mode of research – which could be dubbed research 2.0.

Now, the instantaneous translation of a text, a web page or an email message is just a click away. MT results won't be elegant, but may help users not sharing a common language, or poor bilinguals, to communicate, if users are prepared to put in the extra effort often required to repair poor grammar and wrong word choice.

Ironically, it is translators who provided and keep providing the bi-texts needed to feed the SMT engines. The faster translators generate bi-texts, the better the engines' output becomes..

This freely available, 'unassisted' MT should, however, not worry most professional translators much. Its counterpart, 'assisted' MT, should. When engines are tailored to specific needs with specific glossaries and bilingual data, MT produces better results. When technical writers craft the source text under simplified English/controlled language principles, automatic translation output improves.

Then, this output may or may not be assisted by way of post-editing. If no post-editing is required, 'assisted' MT often takes jobs formerly performed by professional translators. If post-editing is needed, it morphs translators into proofreaders of MT.

The first case is best illustrated by how corporations now deal with their knowledge bases. These hold thousands of articles mostly related to troubleshooting and customer support, that are mostly used by IT experts, most of whom would have at least some rudimentary knowledge of English. In previous years and depending on web-page traffic, the most requested articles would have been translated by professional translators into the most requested languages. The trend now is to provide only MT support.

Internal research, such as that reported in *ClientSide News* (Dillinger & Gerber, 2009), seems to indicate that the level of acceptance of articles translated with MT, as measured by user feedback, is practically indistinguishable from the level of acceptance of articles that are professionally translated. Thus, the unwillingness of clients to commission further translations is solidified.

Raw MT output, however, is not yet suitable when web content is addressed to general users with no command of English and no IT expertise. For these, human mediation is still required. Corporations are increasingly dealing with this in a new way, however. Apart from providing translators with memory and terminology, they will require 'no match' segments to be fed with MT output. Translators working on them no longer do so from scratch, but from an MT baseline.

This integration of MT with translation memory (we could call it MT-assisted TM), already attempted without much success in the nineties when MT came on a CD-ROM, seems to work well now with more powerful engines accessed online. In the process, post-editing, in the past linked to a quick fixing of MT output to make it useful for in-bound (assimilation) purposes, has gained a new meaning: it is now another approach towards full translation, translation for out-bound (dissemination) purposes.

The Google Translator Toolkit, released in June 2009, is a good example of this MT-assisted TM: if no matches are available in the memory, the Toolkit advises its users, to fill in the target segments with MT. In the continuum line, with human translation at one end and machine translation at the other, we seem to be leaving the machine-assisted human translation stage behind, to move into human-assisted machine translation. TM, which branched out of MT research around 1980, with Kay's (1980/1997) seminal paper as a signpost, branches back into MT as we speak.

As with manufacturing, the greater the role machines play, the less skill is needed to operate them. Machines facilitate the access of semi-professionals and amateurs, to tasks previously reserved for professionals. What was also new about the Translator Toolkit was that for the first time technology (memories and terminology bases, plus the engines to leverage them) previously reserved for the professional translator, was now made available to amateurs. Google launched the Toolkit, one assumes, in order to gather bilingual data to feed its Translate engine. Make the web accessible across languages is what matters in the Google's big picture, not to become a player in the small translation industry market. As an unintended consequence, however, amateur translators have now another toy to play with.

We will now move on the other trend disrupting professional translation on the web: the irruption of the non-professional.

4. The Charybdis – crowdsourcing

Crowdsourcing became a buzzword following the success of Howe's (2008) book. It entered into the translation profession when Facebook did away with language service providers and professional translators, and started localising its site by using the skills of those who better knew Facebook – its

(bilingual) users. This was not driven by a desire to cut costs. Creating the platform to enter the contributions of volunteers where users could vote on them, and then implementing the changes must not have been cheap. However, it worked well as a community building exercise, and a perusal of the Spanish (Spain) version indicates to me, that it could not have improved much had Facebook used professionals. Criticisms by translators (initial occurrences of *aser* instead of *hacer* in Spanish) did not hold weight and errors were corrected. A new strategy for quality assurance emerged, based not on the opinion of the expert, but on votes, on the wisdom of crowds (to make a reference to Surowiecki's 2004 book).

A second collision with crowdsourcing involved LinkedIn imitating Facebook with a twist: asking those trained translators amongst its users to offer their services for free. It backfired, as was timely reported by CommonSense Advisory's Kelly (2009).

Crowdsourcing is now central to industry concerns. The Translation Automation User Society (TAUS) theorises it as community translation, and CommonSense Advisory as C3 (community translation, collaborative technology and crowdsourcing). Under the radar, corporations are turning back the clock to the pre-nineties when translation was done in-house by bilingual employees. What is now different is that those educated, keen bilinguals on the corporation payroll now have better support. Web-based technology offers them, apart from the generic search facilities, the advantage of easy access to massive bilingual databases (such as TAUS Data Association, MyMemory or Linguee). What is also different is that, in the fluid web environment, translation errors don't carry big risks if fixed quickly. The trend cannot but progress with new tools emerging that incorporate crowd support (crowdin.net, for example).

Despite the fact that content on the web is growing exponentially, the pie for paid professional translation seems to be shrinking, and the price per word pushing downwards. This observation is based on anecdotal evidence (entries on professional lists, comments by colleagues), and seems to apply particularly to first world countries and to languages of greater demand. We are immersed in an economic crisis, but translators tend to blame technological advances far more than the crisis.

It would be good to find ways to quantify the percentage of translation on the web done by professionals as opposed to that done by volunteers. Professionals can only be paid by commercial and institutional interests. The FOSS community has maintained a volunteer approach to translation matters since its inception – it did not have a choice. There is little money in translating most content related to cultural, social and political activism. Even less for translating in the area of user-generated content, from blogging to instant messaging.

Professional translators can take the moral high ground and claim the quality of non-professional translation is low, pointing to ethical or even legal implications that are overlooked, but that will do nothing to reverse the process. A process that, overall, may well be good for society, if not for the profession.

The crowdsourcing trend shows that the web values translation as a skill, if not as a profession. After all, translation is a skill, in the same way that writing is a skill, rather than a profession. The ability to write is the mark of the educated person. The ability to translate is the mark of the educated bilingual. Translating is just the fifth macro skill, alongside speaking, listening, reading and writing, as Campbell (2002) put it. Mass amateurisation as it applies to translation is a sign of social health, as literacy is. There is nothing wrong with the cult of the amateur (Keen, 2007) as it applies to translation.

There is room for professional writing (in literature, in journalism, in engineering), but writing is mostly a highly valued skill relevant to any professional endeavour. Most writing in literate societies is done by subject matter experts, not by professional writers. It is the same with translation. If this means the end of translation as we know it, so be it.

Scribes were not happy when the printing press made what had been a prestigious profession obsolete. Translators now are luckier than scribes were, however. The skill of the scribe had nothing to do with the skill required of the composer of lead types, and once the printing press succeeded, it was irrelevant. In the web age, the skill of the translator, transferring meaning across languages, not just words, will still be highly valued.

5. Scanning the horizon

In this environment, what is there left for professional translators to do? Should they resign themselves to the slow and painful death of their profession? Translators seem to be facing a challenge of Homeric proportions, but let us put things in perspective.

Translation was for centuries one of the many activities that educated bilinguals could do, so long as they had some understanding of the subject matter involved. In specialised areas (such as international relations or Bible translation), some training along the lines of a master-apprenticeship model may have been available, but that was the exception rather than the rule. Translation as a profession based on language-transfer expertise, rather than on subject-matter expertise, developed only in the second half of the twentieth century, when the master-apprenticeship model proved unable to cope with the growing demand of scientific and technical development. It was then that universities started offering translation degrees (and lecturers began developing translation studies as a distinctive discipline).

This movement from the translator as a subject-matter expert to the translator as a language-transfer expert (and back), also took place in the shorter cycle specifically linked to computing. In the pre-nineties era, translation was done in-house and by bilingual employees – on today's parlance we could say translation was then crowdsourced. In the nineties, this practice was seen to slow the pace of development, translation needs were outsourced, and the translation industry developed in earnest.

Then the arrival of the web changed the rules of the game, not just for translators, but for all professions. The one-to-many models of communication in print, broadcast radio and television, gave way to the many-to-many model of the web. On top of that, the web could embed all previous forms, including the one-to-one (the phone call, the letter). Shirky (2008) presents an interesting account of this. Traditionally, one-to-many models were based on institutions, clear role delimitation, professionals as gate-keepers, setting as high an entry threshold as the insider discourse of norms and standards could allow. The many-to-many model of the web favours collaboration. Anyone and everyone can set up shop and become a writer, teacher or adviser, if they are eloquent enough to find an audience who is ready to listen, to be taught or to follow advice.

This mass amateurisation will have little impact on unskilled, manual labour. Garbage collectors fill a most important social function, and it's unlikely that the web will threaten their working conditions, even if it allows for the rising up of communities of volunteers to clean up a river or a park once a year. Mass amateurisation will also be limited in highly regulated professions such as law, medicine or civil engineering – although the web may facilitate the self-writing of wills and contracts, and even the self-diagnosis of real or imaginary illnesses. In the realm, however, of vocational, less regulated professions, such as journalism, photography or translation, the web has dramatically lowered the entry threshold.

Besides that, translators now have to contend with machine translation. Despite advances in machine learning, programs that with minimal human input write laws, medical vademecums or even computer help files, are still in the far future. On the other hand, programs that can translate texts in a useful way for some purposes are already available.

Machines able to perform tasks that previously required the toiling effort of thousands have liberated this

manpower for, hopefully, more interesting ends. In the same way earth-moving machinery can do in hours the work that took hundreds of workers months to complete, machine translation, effortlessly crunching words by the million, can liberate translators' time for more skilled matters. There are situations in which earthmoving machinery will be unsuitable – on a paleontological camp site, for example. In most cases, a human touch will be needed to complete the work of the machine. It is the same for translation.

Here we may comfort distressed Luddites by referring to the so called lump of labour fallacy. It is not that the amount of labour remains constant and machines condemn humans to unemployment. The time saved by machines will be employed in tasks requiring other skills, including the skills involved in the development, operation and maintenance of those machines. Technological advances seem to create a world in which the costs involved in the production and sale of goods and services grow in the areas of research and development and decline in manufacturing. What is most valued now is the ability to create, to improve, to express, abilities that translators are constantly developing by default as they work – across two languages. It is just a matter of being strategic when placing oneself in the labour market.

6. Charting the course towards professional security

The era of the professional translator as a language-transfer expert is nearing its end. Translation as a skill – which, as with all skills involved in writing, takes a long time to develop – is on the rise; translation as a profession is not. With language-transfer skills alone, sooner rather than later, the professional will collide with non-professionals taking their jobs or with seeing those jobs disappear in the machine translation swirl.

Tasks that required the professional translator's keyboard in 2001 may not require it in 2010. As the pendulum swings back to the time of the content-matter expert that can translate, it pays for the professional language-transfer expert of today, to do some SWOT-ing on the issue, and to make the strategic decision of either embracing or avoiding the brave new world of Web 2.0.

Avoiding it means finding a niche, specialising, becoming a subject-matter expert and working in those highly regulated areas in which text granularity does not yet suit the grinders of word-crunching machine translation, where the risks of getting things wrong compel clients to avoid the temptation of reaching for the keen amateur.

Embracing it, means understanding professional translation as a hub and the translator's role as that of a linguistic consultant and quality assurance (QA) expert, advising the client for which task and at which point translating from the machine translation baseline and/or involving non professionals may or may not be advantageous. Melby, whose 'translator's workstation' envisaged in the early eighties what was to become TM, in parallel with Kay's 'translator's amanuensis', referred already to this hub strategy in these pages of *Tradumatica* in 2006: by... about now, he wrote, «the only kind of non-literary translator who will be in demand [will be] one who can craft coherent target texts that, when appropriate, override the blind suggestions of the computer» (Melby, 2006), with the well-paid translators being those involved in the entire QA process.

This hub approach may even go beyond providing QA and linguistic consultancy. I can imagine some translators soon applying their cross-cultural skills to craft coherent *source* texts that MT can render, usefully if not elegantly, into the target language(s), pre-editing this source on next generation tradukka.com-like interfaces. So, the translator may write in Spanish the text for the web site of a small tourist resort in a way that it can (machine) translate itself into English (and French, and German) without loss of functionality.

The rules of web navigation will keep changing. Research is needed to bring out the new patterns, and to inform translation training and professional development. Then, as it was in the time of Homer's Odyssey,

talent will help the luckier of us chart the course towards professional security.

References

- Campbell, Stuart (2002) "Translation in the Context of EFL - The Fifth Macroskill?" *TEFLIN*, 13, 1.
- Dillinger, Mike, and Gerber, Laurie (2009). "Success with Machine Translation. Automating Knowledge-base Translation". *ClientSide News*, January, 10-11.
- Howe, Jeff (2004) *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*. New York: Crown.
- Kay, Martin (1997). "The Proper Place of Men and Machines in Language Translation", *Machine Translation* 12, 1-2, 3-23.
- Keen, Andrew (2007). *The Cult of the Amateur. How blogs, MySpace, YouTube, and the rest of today's user-generated media are destroying our economy, our culture, and our values*. New York: Doubleday
- Kelly, Nataly (2009) Freelance translators clash with LinkedIn over crowdsourced translations. From <http://www.globalwatchtower.com/2009/06/19/linkedin-ct3/>
- Melby, Alan K. (2006). "MT+TM+QA: The Future is Ours". *Tradumatica* 4, 1-6. From <http://www.fti.uab.es/tradumatica/revista/num4/articles/04/04art.htm>.
- Shirky, Clay (2008). *Here Comes Everybody: The Power of Organizing Without Organizations*. New York: Penguin.
- Surowiecki, James (2008). *The Wisdom of Crowds. Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations*. New York: Doubleday.