Article

Crowdsourcing as Human-Machine Translation (HMT)

Dimitra Anastasiou

SFB/TR8 Spatial Cognition
Computer Science & Languages and Literary Studies
University of Bremen, Germany

Rajat Gupta

Centre for Next Generation Localisation, Localisation Research Centre, Department of Computer Science and
Information Systems, Limerick, Ireland

Abstract

In this paper we examine the model of crowdsourcing for translation and compare it to Machine Translation (MT). The large volume of material to be translated, the translation of this material into many languages combined with tight deadlines lead enterprises today to follow either crowdsourcing or MT. We see that crowdsourcing tends to be the human "version" of MT nowadays, as they share some common characteristics: they can cope with high volume, perform at high speed, and reduce the cost of translation. MT is an older technology, whereas crowdsourcing is a relatively new phenomenon gaining much ground over time, mainly through Web 2.0 features. Examples and challenges of both models will be discussed and the paper is closed with future prospects regarding the combination of crowdsourcing and MT, so that they are not regarded as opponents, but rather as components. These prospects are partially based on the results of the survey we conducted.

1. Introduction

Crowdsourcing is a new model that opens the way for collaborative processes; the focus is on the way users and companies interact through the medium of the Internet. It is not any longer about an employee's obligation to carry out an assignment by an employer, but about a large group of people who are available and willing to perform a task that an outsourcer(s) asked for. It should be pointed out that crowdsourcing can also be used outside the companies as well.

Crowdsourcing has substantially changed business models, as customers and users become potential partners. There is a shift from buying, selling, and consuming to participating in the whole workflow. This is the crowdsourcing situation today; however, it can be changed, as crowdsourcing is a trend indeed and thus constantly adapted to time's and market's needs and preferences; crowdsourcing is not a technology from the view that it does not need an engine (as in the case of MT), but is facilitated through Web 2.0 technology.

In this paper we focus on translation and specifically on the relatively new trend of human/community-based translation: crowdsourcing (see Section 2); we also focus on the quite older technology of Machine Translation (MT) (see Section 3). In 2009 the pioneer of crowdsourcing, Jeff Howe, mentioned at a forum [1] that "[MT] didn't fit into the taxonomy of crowdsourcing [he]'d created" following with "I couldn't figure out where to put it". In our attempt to fill this gap, we investigate in section 4 the common points and differences between the two models, crowdsourcing and MT, and conclude with our viewpoint of crowdsourcing as human MT (HMT). Section 5 focuses on the survey we conducted regarding crowdsourcing and its relation to MT, and in section 6 we refer to some future prospects combining the two models.

2. Crowdsourcing

In this section we introduce crowdsourcing starting with its definition by the pioneer Howe; we continue with our suggested redefinition, and furnish some examples of crowdsourcing's implementation.

The term "crowdsourcing" was first coined by Jeff Howe [2] in his *Wired* Magazine article "The Rise of Crowdsourcing", while in his blog the same month he defined it as follows:

"Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call." [3].

It is different from an old, typical work delegation, where an employer asks an employee to undertake a task. In our opinion, what is missing from the aforementioned definition is that the reward of performing the task is not high, if any at all. In terms of translation, crowdsourcing in most cases brings no income to translators. This is mainly because supply is greater than demand. In other words, there are many translators who are available and willing to translate something without getting paid and the content is divided among them.

In addition, we disagree on the point that the "crowd" is "undefined and generally large". Very often the people performing the task are very carefully selected and managed (see Expert Decision, Spigit[4]). According to this Expert Decision model, designated experts associated with the sponsoring organisation review the submissions, identifying those best addressing the objective of the contest. The sentiment of the crowd is not an element in these contests, as they typically address more technical challenges.

Here we would like to distinguish between crowdsourcing and community-based platforms; two terms often used interchangeably which is incorrect. Henceforth we will refer to *crowdsourcing translation*. By crowdsourcing translation we mean human translation using crowdsourcing platforms (see 2.1). There are different alternative terms established, such as crowd translation, collaborative translation, community translation, social translation, hive translation, etc., but we prefer the term *crowdsourcing translation*, because it makes the crowdsourcing phenomenon explicit. In *crowdsourcing translation* terms, on the one hand the *community* translators are connected with each other and they typically have a vested interest in the content. They are a better defined group of people with high passion, dedication, emotional attachment, and are driven by a desire to be involved and actively participate. On the other hand, generally a *crowd* is undefined, uncontrolled, and can lead to crowdslapping (see [5]). In crowdsourcing translation, there is no inherent connection between *crowd* translators and no connection to the content either. It should be pointed out though that communities are created out of the crowd; so the importance of the crowd is undeniable. A relevant article pertaining to the differences between community and crowd can be found in [6]. In addition, [7] compares many, unconnected individuals with a committed set of connected members resulting to 'lightweight', i.e. weak-tie vs. 'heavyweight', strong-tie attachment, respectively. This attachment is result of various dimensions, such as task interdependence, authority control, and group focus.

Crowdsourcing is often regarded as a buzz word, but we believe that it is much more than just hype in that: it is a successful productivity model and its adoption by many organisations proves that; examples will follow in the next paragraphs. In our attempt to redefine crowdsourcing, we would like to replace the "undefined, generally large group of people" of [3]'s definition with "community", as we believe that it is not the crowd that contributes, but only a dedicated part of it. Moreover, we focus on community's wisdom, as Jeff Howe points out in his presentation [8]: "The person who you think is best qualified to perform a job is not always the best person to do it"; the wisdom is out there, spread throughout the world and crowdsourcing makes it possible to "tap into this wisdom". The profit that crowdsourcing brings to organisations should not be neglected either; organisations want to increase their return on investment (ROI), and cope with tight deadlines and multilingual simultaneous shipping. Based on all the aforementioned points we redefine crowdsourcing as follows:

"Crowdsourcing is the process by means of which organisations can tap into the wisdom of their dedicated external community and use the wisdom for their benefit, i.e. with low cost, for more languages, and within the specified time frame."

We adopt this definition as a starting point; the technologies of the World Wide Web are emerging constantly and thus the definition of the "community" changes too; as crowdsourcing is a trend, we are cautious about the validity of this definition even in the near future.

Also, the ROI increase is not the only motivation of companies to do crowdsourcing. There are many not-for-profit organisations which through crowdsourcing contribute to humanitarian work, bridge the digital divide, and many times save lives. To give an example, in January 2010 a group of people were gathered in a part of The Fletcher School and exchanged GPS coordinates with other people being close to the airport for seven key locations across Port-au-Prince where many Haitians were known to be trapped under rubble [9].

Another example is the not-for-profit organisation *The Rosetta Foundation*[1] which is designed to combat lack of access to basic information on health, education, and justice, and remove language barriers by providing an infrastructure for translation and localisation. The Rosetta Foundation caters for translation requirements for underserved customers and goes beyond the scope of mainstream translation services, involving communities on the ground.

It is important to mention that crowdsourcing is not only related with translation, but with any job which is now "outsourced" to the general public. Here we refer to some examples (not related to translation) of successful implementation of crowdsourcing given in Howe' presentation: *Amazon Mechanical Turk, Threadless*, and *iStockphoto*.

- *Amazon Mechanical Turk* is a successful implementation of crowdsourcing which is about breaking down tasks to little Human Intelligence Task (HIT) among the community, such as completing surveys or answering short questions. The Mechanical Turk Requester gives micro payments, though only when he/she is satisfied with the results. [10] state that "the low cost, scalable workforce available through Mechanical Turk (MTurk) and other crowdsourcing sites opens new possibilities for annotating speech and text, and has the potential to dramatically change how we create data for human language technologies".
- *Threadless* is a company utilising the full potential of the crowdsourced ideas in implementing T-shirt designs. The model of Threadless works in such a way that anyone can join the community by a simple process of registration, i.e. providing a working email ID. Then the user can contribute their own designs for T-shirts and the best design or the winner is rewarded with $1,500 cash and a $500 gift voucher from Threadless. Threadless produces T-shirts which are like any other designer T-shirts for as cheap as $10 to $15.
- *iStockphoto* is a company which produces affordable stock photos by collecting royalty-free stock photography through a web portal. The idea behind iStockphoto is to invite photographers – amateur and expert – to participate in the form of an open call, and in return everybody uploads their creative content. iStockphoto then invites all the clients to buy the best stock photographs.

Another example of crowdsourcing implementation that in our opinion shows high creativity and revolution is a car made by "tweets". Thousands of ordinary people with an email address and/or a Twitter account designed the first cyber car called "the Rally Fighter"; *Local Motors* is the company that oversaw its production and the whole process has taken 14 months and about $2m. This project definitely changed the way the car industries work, as *General Motors* has recently started a blog and a "feedback loop" where customers can share their opinions and the Citroën C3 Picasso was designed with the aid of members of the public who drove various prototypes of the car and were filmed by hidden cameras.

Crowdsourcing has revolutionised the way people contribute for the fulfillment of a common goal and Web 2.0 supports that. *Web 2.0* and *crowdsourcing* are terms which are incorrectly interchanged. Web 2.0 tools and technologies facilitate crowdsourcing platforms; in other words, Web 2.0 is the means, the "engine" of crowdsourcing. The term *Web 2.0* was first coined by [11] who foresaw the future of the Web and its dominance even on electric appliances:

"The Web we know now (...) is only an embryo of the Web to come. (…)The Web will be understood (…) as a transport mechanism, the ether through which interactivity happens. It will [...] appear on your computer screen, [...] on your TV set [...] your car dashboard [...] your cell phone [...] hand-held game machines [...] and maybe even your microwave."

In the first Web 2.0 Conference in 2004, Tim O' Reilly described Web 2.0 as the concept behind the "Web as Platform", where software applications are built upon the Web as opposed to upon the desktop. Further description of Web 2.0 is outside the scope of this paper. Besides Web 2.0, there are other collaboration models, such as peer to peer (P2P), Rich site summary (RSS), folksonomy, etc. (see [12]). Today there are workshops organised related to Crowdsourcing and Translation[2] (by Bederon and Resnik), Ubiquitous Crowdsourcing[3], Social Media Engagement[4], and so on. To recapitulate, crowdsourcing is a successful productivity model facilitated through Web 2.0 features where the community's wisdom is used to achieve an organisation's task. In the next subsection 2.1 we present a crowdsourcing translation workflow, some examples of crowdsourcing translation as well as crowdsourcing's advantages (2.2), and challenges (2.3).

---

[1] http://www.therosettafoundation.org/, 05/02/11

[2] http://www.cs.umd.edu/hcil/monotrans/workshop/, 05/02/11

[3] http://www.personal.psu.edu/u1o/crowdsourcing/, 05/02/11

[4] http://wwwhome.math.utwente.nl/~volkovichyv/some2011, 05/02/11
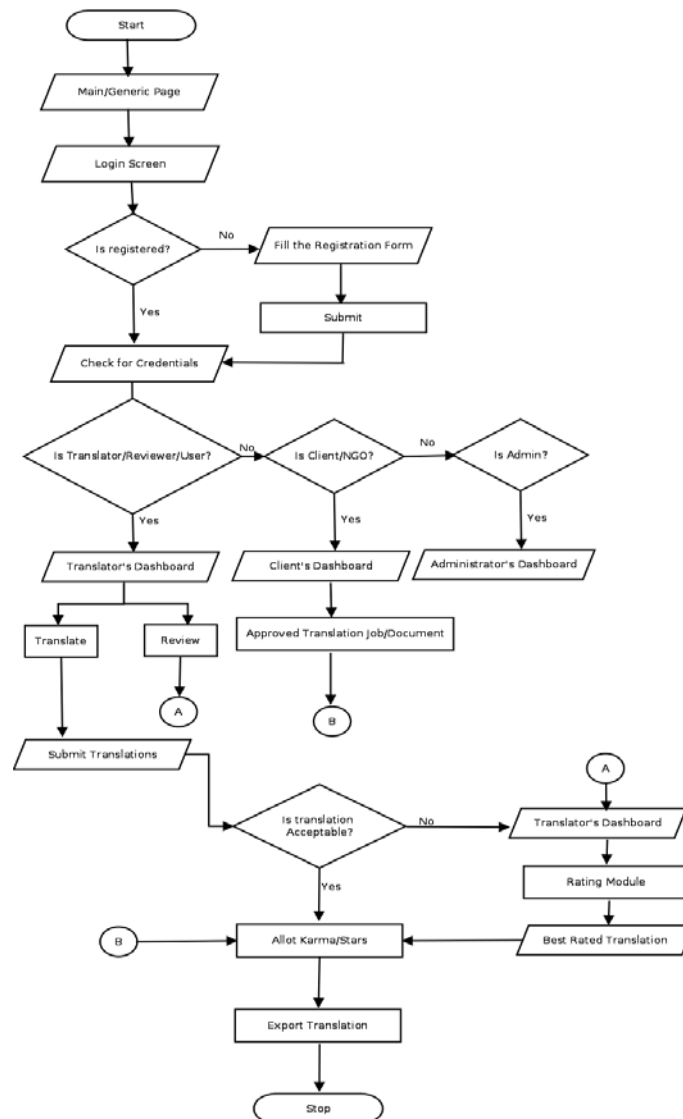
2.1. Crowdsourcing Translation

With crowdsourcing translation we mean the existence of a crowdsourcing platform and a volunteer community/crowd willing to translate with low payment (if any) content which is submitted online. In the next paragraphs we compare a short typical outsourcing translation with a crowdsourcing translation workflow.

On the one hand, a standard outsourcing translation workflow includes in summary the following steps: existence of an offline database of accredited translators categorised according to their language pairs and their rates, a project manager (PM) selects suitable translators for a specific project, and after checking their availability, he/she sends them the content to be translated. Then the PM waits for the translated content to be submitted. After the submission, the PM sends the translated content to proofreaders to be reviewed, and then the content is ready to be released. It should be mentioned that we left out steps pertaining to budget involvement, including preparation, planning, approval, billing and collecting, risk management, and so on. Also, content nowadays needs to be translated in more than one language which makes this workflow repetitive. It is clear how complex, time consuming, and cost intensive such a typical outsourcing translation can be. In addition, communication between translators and PMs is decelerated by many intermediate steps and many ineffective ways of information and data exchange.

On the other hand, a standard crowdsourcing translation workflow includes a website starting with a main page, followed by a log in screen and registration. According to credentials, registered people are categorised to translators/reviewers, clients, or administrators, and are proceeded to the corresponding dashboard. After the translations are submitted, they go for approval and rating, and the best translation is accepted and often allotted karma/stars. "Karma" is a way of showing the contribution of the user and is calculated based on what sort of work the user did and when he/she did it[5]. Figure 1 illustrates an activity chart that was used to prototype the initial test bed crowdsourcing application in our research.

---

[5] https://help.launchpad.net/YourAccount/Karma, 05/02/11

Figure 1. Activity figure for a basic crowdsourcing translation platform

After the comparison of outsourcing with crowdsourcing and our conclusion that the latter is more effective in terms of time-to-market (TTM), cost, and multilingualism, now we refer to some technologies (systems, platforms) which have been translated through crowdsourcing, e.g. Ubuntu, Google, Facebook, klingon, dotSUB, and Worldwide Lexicon. The reasons why users of these and other technologies translate the content are discussed in 2.2 and 2.3.

Although the term *crowdsourcing* was first coined in 2006, translation of open source tools by their users and not professional translators is not a new phenomenon. Linux is maybe the oldest example, which was translated by its own users and still today Ubuntu, a "distribution" of Linux, is translated by emotionally attached users who want to use the systems in their own mother tongue.

Moreover, the service "Google in your own language (GIYL)" [13] was one of the first projects to be translated by volunteer translators and often its users. There are often updated various projects offered and individual users with a Google account can sign up and help translate Google's products into languages Google does not currently offer.

Facebook is perhaps the most recent example of a major platform which has been translated using crowdsourcing. In March 2008, the entire Facebook site was translated into French in a 24-hour period by 4,000 dedicated French speakers [14]. Sun Microsystems, Adobe and other large corporations have also turned to crowdsourcing to translate their content. Crowdsourcing translation is subject to controversy however, and petitions against "unethical business practices" have been signed [15].

Another example of crowdsourcing translation is klingon, a language institute which promotes and supports this unique artificial language and culture of the aliens of Star Trek, and provides a forum for discussion and exchange of ideas.

Moreover, a successful project of crowdsourcing is the "TED Open Translation Project" [16] which offers subtitles, interactive transcripts, and the ability for any talk to be translated by volunteers worldwide. TED and its technology partner, dotSUB, developed a set of tools that allow participants around the world to view, upload, transcribe, and translate any video into and from any language by creating their own subtitles.

Another crowdsourcing/collaborative translation platform is the Worldwide Lexicon (WWL) [17] which combines MT with submissions from volunteers and professional translators. The user downloads the WWL Firefox plugin and can submit translations of web pages. WWL renders any web page in any preferred language combining MT with Translation Memory (TM) matches. A relevant project (and not a platform) is CoSyne [18], Multilingual Content Synchronization with Wikis. In CoSyne the open-source MediaWiki [19] platform is used and translation of user-generated content (UGC) between 6 core languages is achieved. Also corrections of MT output help improve MT performance.

In this section we made an overview of crowdsourcing translation by providing some characteristic examples; in the next two sections we discuss the advantages and challenges of crowdsourcing, particularly in translation environment.

2.2. Advantages

Without a shadow of doubt, crowdsourcing translation is a business model which can be very efficient. In this subsection we examine three different advantages in deploying a crowdsourcing platform from the side of companies:

1. Multilingual support;
2. Quick solution;
3. Monetary benefits;

Crowdsourcing facilitates multilingual support, since – being human/manual translation – it gives to the translators the opportunity to translate from and into any natural language they want; there is no limitation and/or restriction. A report by Common Sense Advisory in March 2011[6] for buyers of language services concludes that crowdsourced translation yields more multilingual content in less time. They state that crowdsourced translation is one way to reach more markets and constituencies: "as more organizations use community translation across various industries, it will become a requested tool to support globalisation. Implemented properly, crowdsourced translation allows organizations to create lasting and meaningful engagement with a larger segment of their customers and constituencies." As for MT regarding multilingual support, MT supports often only some languages (combinations); so the companies which use MT for their translation purposes are restricted only to some languages[7].

Moving to the second advantage (quick solution), we do not consider here the preparation and implementation of crowdsourcing, as this certainly includes project management, planning, strategy, and good attendance. Instead we mean the actual time a task needs to be completed. For example, a small crowdsourcing task may take only some minutes to be carried out, because it is "open" call (see definition of crowdsourcing by [3] in section 2) and many people have access to the task. So the fact that many people and not only one person carries out the task(s), helps accelerating the process. As pointed out by [20], there is big range of talents and in the era of crowdsourcing there are plenty of eye balls and everything can be solved: "Given enough eye balls all bugs are shallow". There is a coordination complexity risk, because the communication between many people is mostly more difficult than between few people; the strategic project management, as mentioned previously, should play a role at this stage. Possibly limiting the task which goes through crowdsourcing and giving higher weight to one or two project manager(s)/crowd's leaders bridge the gap of communication complexity (see chapter 6).

As for the monetary benefits, the company which implements crowdsourcing is the main partner which draws these benefits. The implementation of collaborative platform is a relatively cheap solution for the organisation in comparison to a standard outsourcing workflow. Certainly initially trained staff is needed to implement a crowdsourcing platform, but in the long run, less maintenance is needed, as it is will be on the hands of the crowd/community to participate and evaluate the best method for the defined problem; this saves both money and time in favour of the organisation. Just to mention that in terms of translation services, the in-house translators in a translation company cost indeed less than external freelancer translators; but in crowdsourcing translation context, we

---

[6] http://www.pr.com/press-release/300309, 05/02/11

[7] There are some exceptions of enterprise editions of proprietary MT systems or also the Google Translate toolkit which provide translation from and into many languages (see 3.2).

do not refer to these external freelancer translators, but to those who would translate something for free and are often not professional translators. On the other hand, often the crowdsourced translated technologies have big impact and higher web traffic because they are available in many languages and sometimes they offer applications which are paid in order to run, e.g. sending a Happy Birthday icon gift through Facebook costs $1. Also, many crowdsourcing platforms start charging in order to pay their hosting costs, e.g. iStockphoto charges a quarter to license images.

Last but not least, apart from enterprise benefit, the community gains from crowdsourcing in that they can use the technology in their own mother tongue, and thus their voice can be easily heard and information is spread around the world – that is particularly important for minor languages. Particularly useful is when the translations are done for humanitarian goals (e.g. translation of medicine manuals), and not only for fun or for social reasons. We would like to close with three more benefits of crowdsourcing translation according to [21]:

1. Translation Prosumption (connection of production with consumption): it is the potential audience for the translation that does the translation. It is not the translator who projects a target-oriented model of translation on to an audience, but the audience producing their own self-representation as a target audience.
2. Post-print translation literacy: translation is linked with prevailing paradigms of literacy, then, as the latter changes, translation in nature is expected to change as well. As we move today from a technological world defined by the printing press to one defined by the electronic computer, reading practices and literacy norms change.
3. Translation and Pluri-subjectivity: In crowdsourcing initiatives we witness reinvestment of translation technology by the human, a strategic use of technical resources to further human concerns or agendas; there is a pluri-subjectivity of interaction.

2.3. Challenges

In this section we refer to the challenges of crowdsourcing, including its limitations or anything difficult and/or negative to cope with. Starting with some technological limitations of crowdsourcing, we then focus on the challenges of quality and spam, motivation, control, leadership.

The Internet has in fact rendered geographical barriers irrelevant, but there are still many technological boundaries today. Crowdsourcing makes sense only in relation to the Internet, something which is not available everywhere in the world. Thus, the participating community consists of the computer savvy class; the typical web user is well educated from the upper or middle class and has good network connections. Apart from that, the time zone barriers still remain and crowdsourcers[8] should bear that in mind when posting a job and expecting its delivery. Information whether distance matters in distributed development and its impact on collaboration can be found in [22].

Quality has always been a controversial topic when people talk about crowdsourcing. Dion Wiggins, CEO of Asia Online, mentioned in his speech at the Localisation Research Centre 2010 that before evaluating quality one must establish quality goals in order to define it: a) define the purpose and b) determine the appropriate quality level. Thus he regards quality as "a factor of purpose". In Wikipedia, specifically, [23] examined how coordination methods affect article quality. They distinguished between explicit coordination (editors plan the article through communication) and implicit coordination (a subset of editors set direction by doing the majority of the work). They concluded that adding more editors to an article improved article quality only when they used appropriate coordination techniques and was harmful when they did not. Implicit coordination through concentrating on the work was more helpful than explicit. Both types of coordination improved quality more when an article was in a formative stage.

In crowdsourcing translation, on one hand, the translation is of good quality, because it is human translation, but on the other hand, there are many disadvantages. In our opinion, the most important disadvantage is the non-professionalism of translation: the community which translates does not often consist of accredited, certified, and experienced translators. Tied with this "unprofessionalism", translators are also not used to using linguistic assets, such as Translation Memories (TMs) and/or Terminology Databases (TDBs) in order to leverage the already translated content and terminology. However, it should be pointed out that crowdsourcing translation does not exclude professional translators from participation; it just additionally gives the opportunity to non-professional, casual translators – often users of the applications to translate. The doors of crowdsourcing are widely open for all and it is on the hands of everybody to enter. A database of expert and casual translators can be created within a crowdsourcing platform as a solution to this challenge. More information about that can be found in section 6.

Another aspect which is interrelated with quality is spamming. [24] made a survey about spams in Amazon Mechanical Turk. Some examples of spams are "Like my YouTube video", "Download this app" and so on. He picked the requesters that joined the market in September 2010 and October 2010 and concluded that 40.92% were marked as spam HITs, 31.83% of which post only spam HITs, while the rest spam and legitimate ones.

---

[8] Crowdsourcer is the person who gives the task to be performed to the public.

Motivation is another very important challenge of crowdsourcing, because if it is missing, crowdsourcing is not feasible. Gaining the attention of people, making them emotionally attached, and motivating them in general to perform the task requested is a difficult achievement. After getting the crowd to perform a task, their actual enthusiasm, passion, and true involvement is what really matters in order to maintain a dedicated group of people, i.e. community, which will be willing to perform another task in the future. The motivation of the crowd can vary: seeking knowledge, recognition and reputation, charity work, money, emotional attachment, and fun. In most cases, in our opinion, the incentive is psychological and not financial; however there are often small payments in terms of crowdsourcing which in most cases are based on the satisfaction of the organisation with the results. Often there are rewards for those who come up with the best ideas and then win cash prizes. A study at HP Labs [25] revealed how attention is related to crowdsourcing activities. On the one side, [25] analysed a massive YouTube dataset which revealed that the productivity of those uploading videos depends on attention measured by the number of downloads, whereas a lack of attention leads to a decrease in the number of videos uploaded. He states that "contributors treat the digital commons as a private good, in which payment for their efforts is in the form of the attention their content gathers." More information about motivation in crowdsourcing can be found in [26] and specifically in Wikipedia in [27]. In addition to motivation and attention, the participation equality is an issue that has to be addressed. [28] stated that 90% are 'lurkers' who never contribute, whereas 9% contribute a little and 1% of users account for all the action. Furthermore, user's voting behaviour has been examined in a quantitative analysis of three online communities (IMDB, Amazon, and BookCrossings) to check whether the crowd is biased by [29]. They concluded that websites with higher barrier to vote introduce a relatively high number of one-off voters, and they appear to attract mostly experts. We also find that one-off voters tend to vote on popular items, while experts mostly vote for obscure, low-rated items.

What follows the challenge of public motivation is the public management and control. A result of high motivation is maybe that somebody tries to control and lead the rest of the crowd. If there is no specific strategy and sequential workflow for performing the task, crowdsourcing can lead to leadership and a subtle, unorganised, and unstructured result. This kind of leadership can be tricky, as the crowd may lose interest in the task as well resulting in a vicious circle. Thus there is need for moderation: somebody in-house to manage the community, and somebody of the community to manage the others in the community; that should not lead to a dictatorial hierarchy, but to a fruitful collaboration. Good control leads to successful collaboration and high quality. The impact coordination and control techniques can have on quality can be found in [23].

Crowd control can be very time consuming and cost intensive, as Biewald of Crowdflower[9] states that many companies do not have the resources to describe tasks, put them up online, and manage the data they get. Nor do they have time for extensive quality-control measures, which include creating "test" questions whose answers are already known, checking responses against one another, tracking individual answer histories, and creating a confidence measure with which to weight the resulting data [30].

3. Machine Translation

Apart from crowdsourcing, Machine Translation (MT) is a technology which increases productivity and thus many organisations use it to (pre)translate their content; some of them would not survive without MT given the high volume and the short time frame of the translated content's release. MT uses computer software to translate text or speech from one natural language (source language – SL) to another language (target language – TL) and its different architectures are rule-based MT (RBMT), statistical MT (SMT), example-based MT (EBMT), and hybrid. A detailed analysis of these architectures is outside the scope of this paper, but we give some examples in subsection 3.2 and some future prospects in 3.4.

In the next subsection 3.1 we take a brief look back to when and how MT started and move towards current MT systems, which are available for free (3.2). Subsection 3.3 presents challenges of MT in relation to speed, cost, and quality, and 3.4 is concerned with a general outlook of MT.

3.1. Machine Translation over years

The beginning of MT may be dated to the mid-1930s, when the first term regarding MT systems was introduced: "translating machines". This term was firstly introduced by French-Armenian Georges Artsrouni and Russian Petr Petrovich Troyanskii (see [31],[32]). However, [33] in a report on scientific and technical dictionaries submitted to Unesco in 1951, stated that "the resulting literary style would be atrocious and fuller of 'howlers' and false values than the worst that any human translator produces". The reason he gave for this statement was the following:

---

[9] http://crowdflower.com/, 06/03/11

"Translation is an art; something which at every step involves personal choice between uncodifiable alternatives; not merely direct substitutions of equated sets of symbols but choices of values dependent for their soundness on the whole antecedent education and personality of the translator".

He describes MT process as "direct substitutions of equated sets of symbols" which cannot surpass the human translation process which includes according education, personality, and thorough thoughts about choosing the best translation candidate.

In May 1951, Bar-Hillel was appointed to conduct research at the Massachusetts Institute of Technology (MIT) and in the next years he collaborated with IBM on a project that resulted in the first public demonstration of an MT system on January 7, 1954. This was a collaboration between Peter Sheridan of IBM and Paul Garvin at Georgetown University. Although a very restricted vocabulary of approximately 250 words and a restricted grammar were used, many MT projects were funded in the USA and the MT research throughout the world commenced.

Then some commercial MT systems appeared, mainly to help translating war-related documentation. The company SYSTRAN was founded by Peter Toma in the USA in 1958, mainly to translate Russian technical documents into English breaking war codes. Also, during the Vietnam War the Logos RBMT system was originally developed in the 1970s to translate English to Vietnamese documentation for the US military. SYSTRAN is still a very powerful MT system, covering nowadays 52 commercially available language pairs and 39 language pairs with limitied availability with the SYSTRAN Enterprise Server 7 edition [34]. In 1960s it became widely assumed that the goal of MT must be the development of fully automatic systems supported by various theories and methods in order to produce high quality translations. There was virtually no serious consideration of how "less than perfect" MT could be used effectively and economically in practice; the researchers seriously believed in Fully Automatic High Quality Translation (FAHQT), the use of human assistance was not seriously considered, and the "pure" theoretical research towards the improvement of MT methods was neglected.

In 1966 it was the ALPAC (Automatic Language Processing Advisory Committee) report [35] which concluded that MT was more expensive, less accurate, and slower than human translators. The decade following the publishing of the report was quiet in the USA and more powerful in Europe, Canada, and Japan. It was in 1982 when Martin Kay [36] proposed Human-Aided MT (HAMT); he defined it as follows:

"By human-aided machine translation, we refer to systems, in which the machine, while retaining the initiative, works with a human consultant, who need not be a translator."

From the point where MT had no human intervention until the final revision, now the human consultant is called upon to resolve otherwise irresolvable problems, but at an unknown cost [36]. It was then in 1997, when Kay proposed *The Translator's Amanuensis*, the early version of our known TMs which could function in the following way:

"[T]he translator might start by issuing a command causing the system to display anything in the store that might be relevant to the text to be translated (...). Before going on, he can examine past and future fragments of text that contain similar material." [37]

To recapitulate, MT started powerfully in the USA, and then later in Europe and Japan. Commercial MT systems were used for a meaningful reason, mainly in war periods to uncodify code. The researchers in the early years neglected the pure theoretical research and the human intervention, something that changed over the years. Because MT is a groundbreaking model with quality that still cannot reach the human translation's quality levels, there were many ups and downs in the history of MT. The earlier systems, like SYSTRAN were rule-based, while the last decade the tendency is statistical MT (SMT) and hybrid models. This started with the work at IBM [38] and large programs that helped MT progress (DARPA, TIDES/GALE, Euromatrix).

In 2006 Jaap van der Meer [39], director of the Translation Automation User Society (TAUS), introduced in his invited keynote at the European Association for Machine Translation (EAMT) conference the Fully Automatic Useful Translation (FAUT) term which speaks for useful – for the industries – MT, not necessarily accurate or of high quality. He states that academics still argue about the most suitable approach to generate FAHQT, while the users start to realise that a mix of the translation technologies today produce FAUT. He voices his opinion that computer power is nearly endless and new technological approaches allow companies to use massive text corpora to train MT systems. We would like to close this subsection with the use of MT today as [40] categorises it in his "Compendium of Translation Software":

1. *Home* use: an automatic translation system designed by the producer for personal use by the general public, i.e. by persons not normally with any experience or training in translation;
2. *Professional* use: system designed for use by professional translators, usually working independently or for translation agencies;

3. *Enterprise* use: system designed for company intranets (e.g. client-server systems) to support a team of translators (often professionally trained);
4. Use *for websites*: system designed for company websites to provide online translation of webpages;
5. *Mobile* use: system designed for use with PDA, hand-held, mobile (phone) devices;
6. *Spoken language* use: system designed to translate speech directly (i.e. not via a text-to-text system).

In the next subsection we focus on the home, professional, and websites' use particularly from the point of open/free MT. The other uses of MT are equally important, but are outside the scope of this paper.

3.2. Open Machine Translation

There is a plethora of MT systems nowadays, both proprietary and open-source. A good distinction between these two can be found in [38]. Regarding the proprietary systems, the first ones were very costly, as there were few out on the market and not many companies and institutions could afford them. Over the years though, the big number of systems lowered the cost and thus MT was adopted more frequently in the business world. Some of the commercial MT systems today are SYSTRAN (earlier RBMT – now hybrid), PROMT (earlier RBMT – now hybrid), Language Weaver (SMT), AppTek TranSphere® (RBMT) and so on.

As for the MT systems available on the Web, [41] points out that there are MT systems that may be freely used, demonstration versions of commercial versions, and non-commercial freely available systems. Based on this distinction and slightly adapting it, we focus here on three different categories of "open" MT systems, that means that the user can use the MT system for free:

1. Open-source MT systems;
2. Online translators of proprietary MT systems;
3. Online MT services.

In the next paragraphs we furnish some examples of systems belonging to each category. As for the first category, the most well known MT open-source systems are RBMT system *Apertium*, SMT – more precisely decoder – *Moses*, *OpenLogos*, EBMT system *Cunei*, and the recent open-source MT system *OpenMaTrEx*. Also, the BLEU metric [42] which measures and benchmarks the MT output is categorised among the open-source MT tools. Furthermore, there are more tools belonging to the general open-source MT software, such as aligners, like Giza++ [43], and language analysers.

OpenMaTrEx is an EBMT system based on the marker hypothesis; it comprises a marker-driven chunker, chunk aligners, and two engines, i) the simple proof-of-concept monotone recombinator *Marclator*[10] and ii) a *Moses*-based decoder [44]. According to [45], Moses with ca. 4,000 downloads in the past twelve months must be the most accessed MT software in the world today. Apertium and Moses offer a demo online, while OpenLogos and Cunei do not. As for all open-source tools, one can download either the toolbox or data dependent on the language-pair. According to [41], the advantages of open-source MT systems are increase of "expertise" and language resources, easy building of a new language pair, and independence. The community participation is active in open-source MT system, as they adapt, develop, and extend the already existing resources. However, as concerns OpenLogos specifically, [45] states that the work of building new language pairs requires large financial resources as well as engineer training, and the technology is now out of date and somewhat closed in on itself, he points out its advantage, like its rich ontology.

We agree with [41]'s opinion that a coordinating team should be managed by a code captain (installation, maintenance, and modifications of the engine's code) and a linguistic captain (responsible for the maintenance of the linguistic data). To sum up, the open-source MT systems are a good solution both because they can be freely used, but also because the code is freely available and everybody can contribute. In our opinion, although many language pairs can be added, often minority languages are covered, and major languages are neglected; for example, in Apertium various Iberian tongues are covered, but there is no option of English-French/German MT[11].

To the second category belong the online MT versions of the proprietary MT tools, such as SYSTRAN and PROMT. The user has the possibility to select between text or website translation. Thus this kind of MT systems is for home or websites use (see 3.1). PROMT offers additionally a mobile translator, i.e. owners of mobile phones with WAP support, communicators, or pocket PCs can translate text, messages, or single words at any time. The limitation though in the mobile translator is the maximum length of 200 characters.

The third category consists of online MT services; this category is distinguished from the first in that one cannot download the code or create/train other MT systems, as in the first category. An example of the third category is

---

[10] http://www.openmatrex.org/marclator/, 05/02/11

[11] It should be taken into account that this depends on funding reasons and client requirements too.

Yahoo Babelfish – Translator. Although its technology is based on SYSTRAN, it is not the official "website" translator of SYSTRAN, as in second category. There are specific limitations in the online MT services, for example users are allowed to translate up to 150 words. In the online version of SYSTRAN there is no such limitation and there are more language pair combinations available than in Babelfish.

Google Translate Toolkit is also an example of the third category. Additionally to MT, users can share TMs and glossaries with everyone or specific users. It is noteworthy that Google Translate Toolkit states explicitly that by submitting content through the service, "you grant Google the permission to use your content permanently to promote, improve or offer the Services", something that users should be aware of. However, Google Translate Toolkit provides the data API[12] which allows client applications to access and/or update translation-related data programmatically.

Also, Microsoft® *Bing Translator* offers an online MT service currently supporting 30 languages, also including a beta Haitian Creole translation engine to help with Haitian relief efforts. It should be mentioned that *Bing Translator* uses SYSTRAN for some language pairs. Recently Microsoft announced the availability of the Collaborative Translations, a technology that combines MT with human translation. In addition to the collaborative features, there are interfaces to translate large amounts of data, support for communicating with the service securely via SSL and the addition of "Translate-and-Speak", a text-to-speech functionality.

Another recent system belonging to the third category, coming from Academia this time is the SMT system *Caitra* [46]. More precisely, it is a computer aided translation tool, as the advantage of this system is that it offers to the users the possibility to select the best candidates and postedit the MT output.

To sum up, three categories of open/free MT exist nowadays: open-source, online versions of commercial systems, and online free services. In our opinion, the first category is targeted at developers, the second at potential buyers of the proprietary systems, and the third at the general public; this is why more MT systems fall into the third category.

3.3. Challenges

In this subsection we give statistics from various reports about MT used today and also provide information about how MT is expected to be in the near future.

The answers that someone often gets when asking about MT's potential might be that "the output is better than nothing", "good enough for casual users who want to get the general gist" and so on. We believe that much research has been carried out, enough to contradict such "easy-to-say" statements, and for some small companies MT is the only way to survive.

Traditionally localisation companies want to deliver their product in many languages quickly, with high quality, and at low cost. Thus speed, quality, and cost are the three important milestones which influence the localisation success. As for the enterprise use of MT, very often indeed the companies pick only two of the following advantages of MT: high speed, high quality, and low price, having to victimise one other, as [47] depicts:

Speed

Pick 2
only

Price   Quality
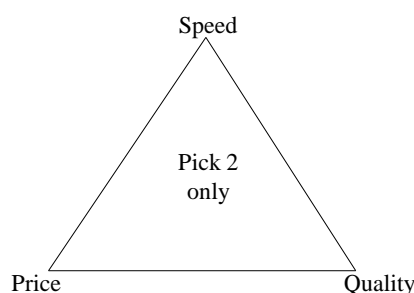
Figure 2. The current localisation paradigm

That means that in the current localisation paradigm one of the three suffers in favour of the other two. The most common case is that MT is often used to save money and time, but this turns against quality. The output quality of MT systems is today still low that makes postediting by human translators necessary. MT quality depends on source language (SL), domain, purpose, and evaluator. The Defense Advanced Research Projects Agency (DARPA) suggests the following as MT quality metrics: adequacy, informativeness, and fluency:

1. "Adequacy" measures how much of the original meaning comes through in the translation;

---

[12] http://code.google.com/apis/gtt/, 05/02/11

2. "Informativeness" is the degree to which information consumers can find what they are looking for and act on what they find;

3. "Fluency" gauges linguistic factors such as spelling and word usage, along with the localization to a given country or market.

Regarding the MT evaluators, they can be information consumers and publishers, or casual users who want to understand foreign-language content. According to [48] access is the top priority for information consumers, while accuracy, linguistic quality, and actionability of the machine-translated content are what matters to the information publishers. Casual users are satisfied with adequacy and informativeness, as they use MT to understand foreign-language content. Noteworthy is that Apptek's MT system scored highest in 2009 NIST Testing of MT systems in both "noisy data" and "informativeness"[13].

As concerns the MT quality improvement, there should be higher investment for high quality proofreading which in turn, takes more time (see section 6). Apart from postediting, there are other ways to enhance MT quality, such as source text control, TM, training, postediting, and general improvements (see [47]):

Improvements

Post-editing

Training

Translation Memory

Source Control

Increasing MT quality
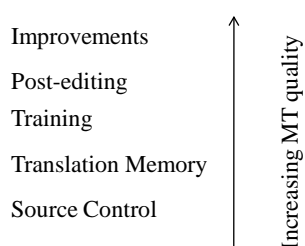
Figure 3. Five factors influence increasing MT quality

In other words, improving MT includes extensive preediting (Controlled Language, Controlled Authoring), cumbersome coding, training, postediting, and specialist teams with careful implementation planning.

As far as speed is concerned, the MT's engine response time is instant (about 12 words/sec). This instant time is relative though, as the MT output is often releasable only after it has been postedited, and as aforementioned, postediting is tied with time consumption. In addition to that, one can argue that a professional translator needs as much time to translate something manually as having it done instantly and correcting it. [49] made a pilot case comparing PROMT and Language Weaver and concluded that a translator's daily output can be produced by a posteditor in less than two hours and that is a 7-fold speed-up for PROMT and a 4-fold speed-up for Language Weaver.

As for the cost related to have an MT system implemented, it should be stated that the cost of an MT system is not equal to the price of the engine, but also of the whole system used, resources, maintenance, training, infrastructure, etc. resulting into a very expensive workflow. Companies should compare this whole workflow to the cost of a translator workstation, consider the training turnaround, the estimated TTM and so on.

To sum up, high quality, high speed, and low cost is the ultimate goal of MT, but often industries have to pick only two out of three options, given the high volume and the release deadline pressure. The third victimised point, which often is quality, can be enhanced through controlled preediting, training, and postediting.

3.4. MT Outlook

Having shed some light to quality, speed, and cost of MT, we now turn our focus to the outlook of MT today and tomorrow. It was in 2009 when the Executive Office of the President and National Economic Council issued its "Strategy for American Innovation" and among the recommendations was a call for "automatic, highly accurate and real-time translation between the major languages of the world – greatly lowering the barriers to international commerce and collaboration". Here we see the importance of MT at a global level in relation to collaboration and commerce, i.e. among others, product, service, and information exchange.

The Common Sense Advisory made a report on the business case for MT which showed that the respondents, owners of deployed MT applications, showed interest and enthusiasm for both the MT technology and its ability to increase the amount of translated information that they can provide to their customers with faster turnaround time and lower costs. [48] mentioned that over half of the non-Anglophone consumers that they surveyed said they use MT when they visit English-language sites. Here we see the prioritisation of use for MT for websites over the other uses mentioned by [40] (see 3.1).

---

13 http://www.itl.nist.gov/iad/mig/tests/mt/2009/ResultsRelease/, 30/06/10

According to another survey conducted by TAUS in 2009, 14% of the localisation service providers (LSPs) answered that they will never use MT, while 40% already use it. Jaap van der Meer at the TAUS User Conference 2009 [50] described the marriage of "data effectiveness" and "profit of sharing", pointing out that the former is hidden in academic circles and the latter is more open in the media and on the Internet. In 2010 TAUS is developing a new service called TAUS Tracker which is an online directory of MT engines providing information about their use cases, business value, and best practices. Tracker users will be able to do quick and refined searches to find MT engines meeting their criteria and reporting will include data at aggregate and granular levels to help users make informed decisions and build MT business cases when necessary.

Another survey [51], the first to analyse the size of the language industry EU-wide, has been carried out in 2009 from the European Commission and particularly the Language Technology Centre (LTC), covering translation, interpreting, localising and globalising, subtitling and dubbing, language technology tools, multilingual conference organisation, and language teaching. According to this survey (see pages 51-67), the use of MT in multilingual workflows is still low. More precisely, it is stated that "562 respondents out of a total of 1103 valid responses refer to language technology tools, but only 36 respondents mention MT, which corresponds to 3.26% of the total responses on tools and 6.41% of all valid responses" (p.52). SYSTRAN was mentioned by 9 respondents, Google translate by 8, Babylon and Language Weaver by 3 respondents each, and Lucy by 2.

As concerns the future prospects of MT, we would like to refer to some current projects and developments related to MT architectures, RBMT, SMT, EBMT, and hybrid. Nowadays there is a tendency of hybrid engines, such as SYSTRAN and PROMT. Both were previously RBMT, but now are based on statistic models too. Today seldom is an MT system purely rule-based, as the first MT systems were; RBMT requires well-trained staff and is very time-consuming. SMT from the other side gains more ground, as it is based on corpora, which become largely available over time. Regarding example-based approaches, EBMT shares many characteristics with TM, as the example database of EBMT can be seen as a kind of TM and the input sentence is matched against this repository. EBMT also shares necessary corpora with SMT. In our opinion, hybrid systems, which combine RBMT with (phrase-based) SMT or EBMT are the systems of the future. As we see from current projects Euromatrix+, MOLTO, and so on, they all employ statistical tools and linguistics rules.

4. Crowdsourcing: Threat to Machine Translation?

In the previous sections we described crowdsourcing and MT separately, but in this section we compare them and try to answer the question whether the former poses a threat to the latter. Although this comparison may seem a bit arbitrary in the beginning, we prove in the end that they share more characteristics than one may think.

The main obvious difference is that crowdsourcing is human translation, while MT is computer automated translation. Thus we first see the differences between human and MT on different points, such as purpose, difficulty of content, friendliness of reader, domain, timeframe, and languages (adapted from [48]):

Table 1. Human vs. Machine

	Human	*Machine*
Purpose	Quality matters	Access is critical
Content Difficulty	Complex	Simple
Friendliness of reader	Any including "hostile"	Friendly or tolerant
Knowledge domain	Any	Predictable
Time frame	Scheduled	Immediate
Languages needed	Budget for just a few	Less budget constraint

From the human translation side, the purpose of the human translation is that quality is the first issue that matters. [52] mentioned that "initial quality [in crowdsourcing translation] is not the top priority. She continues saying that crowdsourcing translation is here to stay and will compete with MT and professional human translation, but it will not take over that whole environment, such as life science or legal translations.

From the MT side, the access to the system is critical, that involves mainly commercial vs. open MT (see subsection 3.2). As for the content, the humans do not have difficulty with complex syntax or semantics, whereas in MT the input text has to be controlled. The domain is highly related with the content; the humans can cope with almost[14] any domain, while most MT systems are domain-specific. This does not mean necessarily that human

---

[14] We say almost because also human translators are specialised in one or more specific domains based on their education, experience, and preferences.

translation is better than MT. Let us exemplify a relevant use case: from one side, we have a human translator to translate a medical document without using any Terminology Database (TDB) and lacking experience in medical translation; from the other side, we have an MT system trained with corpora and terminological references. We believe that the latter brings translation output of higher quality than the former. Moving to proofreading and postediting (see the differences in section 6), commentators/evaluators of human translation are very critical, as they have higher expectations. Posteditors of MT, from the other side, know that the quality is going to be low, so, they are satisfied with any "average" result and have a tolerant attitude. As for the time frame, humans have to submit their translations before the scheduled time frame, while the MT output is instant. Last but not least, when it comes to budget, human translation costs more than MT. Nevertheless, given that MT supports more than one or two language pairs, as normally the human works with, the initial investment to a commercial system can save more money in the long-term. We now replace *Human* with *Crowdsourcing translation* and *Machine* with *Machine Translation* (extended from [53]) and form the following table which outlines the history, the active engine of the model, speed, cost, quality, and profit, as well as other aspects.

Table 2. Crowdsourcing Translation vs. Machine Translation

	Crowdsourcing Translation	*Machine Translation*
Start	2006	1955
Output "engine"	Humans	Computer software
Human involvement	Always	At revision
Control	✘	✔
Terminological consistency	✘	✔
Source text	Uncontrolled	Controlled
Speed	Less than MT	High
Cost	Low implementation cost	Acquisition cost of commercial systems
Quality	High	Low
Profit	Company profits	MT User (single person or company) profits

This table just gives in table form an overview of the year that each model began and also, their basic characteristics. These characteristics are discussed in the next paragraphs.

On one hand, MT is an old technology starting powerfully in 1955 after the Russian into English MT experiment of the IBM-Georgetown University experiment. On the other hand, crowdsourcing is a term coined first in 2006 at the Wired Magazine Blog. Here we do not consider as crowdsourcing translation the translation of open-source tools which is not a new phenomenon.

As for the "machine" which generates the output, it is the humans in crowdsourcing and the computer software in MT. Thus the human involvement is always present through the whole crowdsourcing process and only at the end of the MT process. It is noteworthy that crowdsourcing keeps translators, both professionals and non-professionals, into business; in our opinion, it does not devalue the profession of accredited translators, it just gives the opportunity to more translators to participate (see section 5). However, since crowdsourcing translation is not payed for (or hardly), professional translators will be paid less, as supply is greater than demand.

The control of crowdsourcing translation is a challenge, as there are many people involved, often without a structure, a sequential order. There is no specific time when someone submits something online to be translated and often there is no deadline when the translation should be submitted back. In MT the user has the control in that he/she inputs to the system the content to be translated and in some minute time, it is delivered.

As for the terminological consistency, the use of specific tools and resources is necessary to achieve that and in crowdsourcing it is not often the case, as not many people can either afford or have the time and efficiency to use these tools and technologies. Today there is a trade-off of having professional translators having acquired Computer-aided translation (CAT) systems and non-professionals who lack these tools. In our opinion, it would be preferable to implement leveraging features (TM and TDB) in a free/open source (FOS) MT system and thus have one system in the form of HAMT; then more translators would have access and benefit from the combination of these architectures (see section 6).

Moreover, the quality of the source text strongly affects the language consistency. Often user-generated content (UGC) which concerns personal/community content is crowdsourced for translation. However, the project manager (PM) of the crowdsourcing platform can control the input indeed and post only the suitable content to be translated. But the UGC is so extensive in volume sense, so that it becomes unstructured and difficult to control in practical terms (time constraints). The content to be translated with MT systems can be UGC (which then goes to the uncontrolled structure of the crowdsourcing translation), but very often is used for enterprise content and thus is well authored and controlled (see [54]), and accordingly the output has consistent terminology. According to [55], the use of controlled language (CL) can help save translation costs of up to 40% per language.

We continue with looking at quality and speed, cost and profit from crowdsourcing translation and MT side. The fact that crowdsourcing is human and MT automatic translation affects both quality and speed. Regarding speed, having the same amount of words, the output is generated on-the-fly by an MT system, rather than by crowdsourcing translators, as still there is a workflow to go through in this case (see 2.1). As for quality, by rule human translation is of higher quality than MT, but if one considers the consistent terminology which is element of MT, then MT has also quality values that should not be ignored.

The cost related to the implementation of a crowdsourcing platform seems lower than the implementation of an efficient MT system which ideally covers many language pairs. It needs an initial development of a platform, but then there is not much maintenance and training needed. In crowdsourcing, in general, someone throws a problem to a large audience of vendors, experts, programmers, and volunteers offer solutions and suggestions. [56] refers to some hidden costs of crowdsourcing which are promotion, monitoring and moderating responses, and rating of applications; thus it is questionable whether the use of crowdsourcing is cheaper than a traditional solution. A current example is that SAP solicits ideas for business-friendly social networking applications from a crowdsourcing platform, like Innocentive and they recently launched the Knowledge as a Service (KaaS) cumulusIQ helpdesk. According to Biewald's estimates, the cost for a crowdsourced project ranges from $2,000 to $4,000 for simple tagging projects to $10,000 to $20,000 for more complex custom applications. Stephen Mechler, managing director of the German crowdsourcing Web site *Floxter*[15], calculates that it is 33% less expensive to crowdsource projects like data classification and tagging than to complete them with in-house employees [30].

Accordingly to the cost involvement, the company implementing a crowdsourcing platform profits the most – and also the community at a lower percentage, whereas in case of MT, the one that profits is the user of the MT system who can be either a company or a freelancer translator. Here the advantages of professional editions against limitations of basic free/trial editions should be taken into account.

To recapitulate, crowdsourcing and MT are two separate models with all the implications human and automatic translation involve. Nevertheless, they are both effective business models with high productivity, as both successfully cope with high volume and decrease the cost of translation. Although crowdsourcing translation is not as quick as MT, crowdsourcing translation workflow is still quicker than the typical outsourcing translation workflow. So we can state that both crowdsourcing translation and MT also accelerate the translation process. All these lead to our conclusion that high speed, high volume, and low cost are commonalities of crowdsourcing translation and MT. This is the reason why we consider crowdsourcing as the human version of MT and propose the term human machine translation (HMT). As for quality, crowdsourcing translation has higher quality, because it is human, but MT and TMs benefit from CL, leveraging, and thus consistency. The above can be summarised in the following:

"Crowdsourcing is the human machine translation (HMT) version, as it can achieve translation of high volume, at high speed, and at low cost. The high linguistic quality of human translation and the language consistency of automatic translation can lead to the highest possible translation quality."

The question that arises now is whether crowdsourcing is a threat to MT, that means mainly that enterprises will stop acquiring commercial MT systems and give to the volunteer translators the content to translate. We refer to enterprises because we believe that casual users will continue to access open MT for home and websites use. The reason for thinking that crowdsourcing can be a threat is that crowdsourcing translation shares the advantages of MT, i.e. high volume, high speed, and low cost, and additionally it has the main advantage of higher quality, because it is human translation. According to our opinion, crowdsourcing is not a threat to MT, because although they share some

---

[15] www.froxtel.com, 05/02/11

common characteristics, they still have distinctive differences and qualities, and it is up to the enterprise to select the model that fits their needs depending on the project. We believe in the symbiotic relationship of the two models and more details about this idea can be found in section 6. The survey following in the next section provides some statistics generally about crowdsourcing and also pertaining to its relationship with MT.

5. Survey

In order to evaluate our research findings and make recommendations, we conducted a survey to get feedback regarding motivation, quality (and its improvement), and comparison of crowdsourcing translation with MT. The questionnaire can be found in the Appendix as well online[16].

The respondents were i) 20 participants of the *AsLiB* conference 2009, ii) 27 participants of the *tc world* conference 2009, and iii) 37 students of the modules "Web Applications" and "Machine Translation" at the University of Limerick (spring semester 2009); in total 84 respondents. Most respondents at AsLiB Conference were tools providers, followed by translators, while in tc world localisation service providers (LSPs) were the main respondents. In the next paragraphs we point out some of the survey's results.

To the question if you would translate something for free, the answer "Yes, for a good will" received a response by 30% and 31% by the AsLiB and tc world respondents respectively, while the students preferred more the option "I saw an error and wanted to fix it" (32%). Also, an average of 20% of all three categories of respondents, a noticeable percentage, answered that they would not translate something for free. An interesting remark is that quite high percentage of students would do crowdsourcing translation for recognition and reputation reasons, i.e. to improve their career opportunities and make contacts; this is understandable as they are at young age. Generally speaking, as tastes and temperaments differ with different people, the same is the case with motivation as well; different people have different motivations behind their contribution.
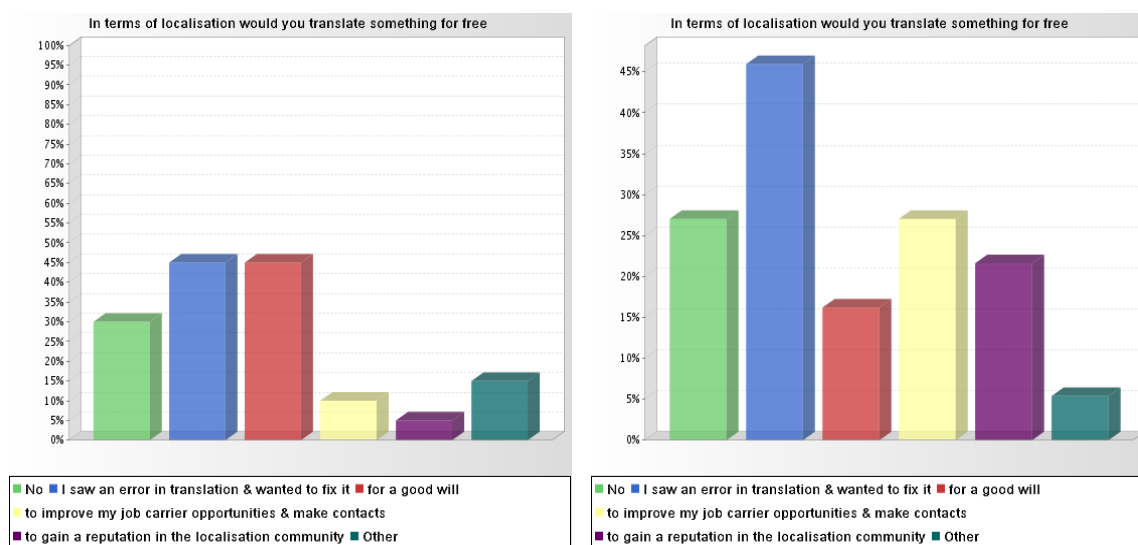
Figure 4. AsLiB's (left) and students' (right) responses on reason of crowdsourcing translation

After the question about translating something for free, we asked the respondents whether they were aware of the concept of crowdsourcing. The average response was "yes" with 78.6%; the ones that were not aware of crowdsourcing were mainly the students (35%).

As concerns the implementation of crowdsourcing platform, we take into account only the participants of the two conferences. AsLiB's predominant answer is "yes" (44%), while most tc world's participants answered "no" (69%). An average of 21% (again excluding students) does not have something implemented yet, but think of doing that in the future.

We come now to a tricky point, the quality of crowdsourcing: taking the average of the whole respondents (three categories), the quality was rated as "good" by 30.3%, as "average" by 29.3%, and as "very good, because it is human translation" by 18%. To the question "How can the quality of crowdsourcing be improved", the answers "through terminology consistency, manual proofreading, administration security control, and checking that the translators are good enough" received approximately equal responses.

---

[16] http://ai.cs.uni-sb.de/~stahl/d-anastasiou/Survey/crowdsourcing_questionnaire.pdf, 29/06/10

Interesting is the next question pertaining to the one who benefits from crowdsourcing. The option "the community" was selected by an average of 41% and "the company" by 33.6%, while "the user" follows with 24%.

The next questions were about the relationship of crowdsourcing with MT which are particularly relevant to the scope of this paper. An average of an overwhelming 81.6% believes that the quality of crowdsourcing translation is higher than that of MT as can be seen in Figure 5.
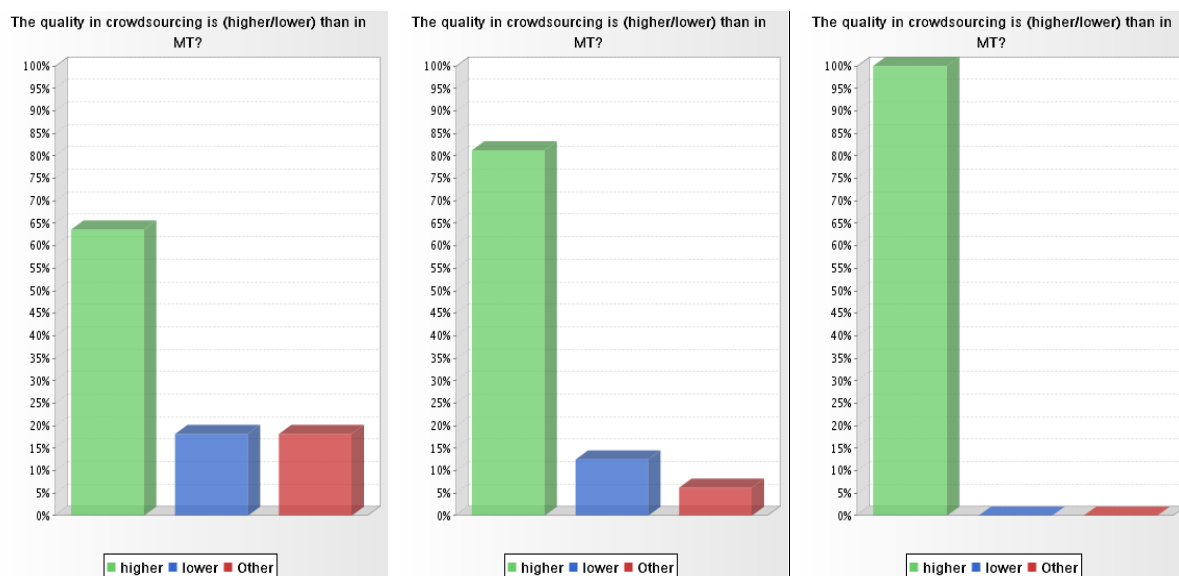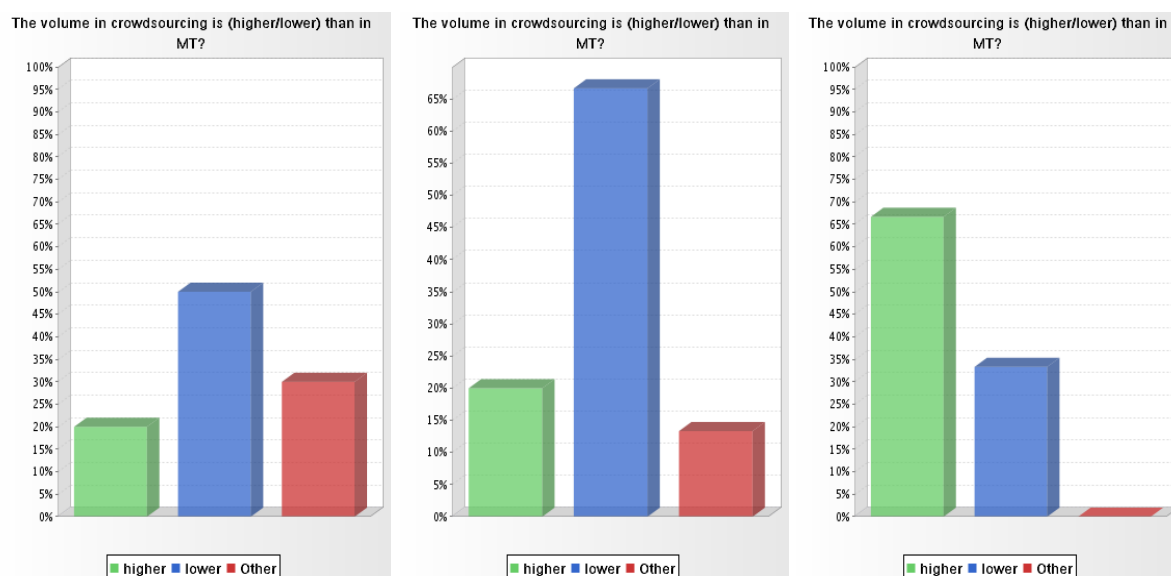

Figure 5. Responses on whether crowdsourcing translation output is of higher/lower quality than MT output (AsLiB – left, tc world – middle, students – right)

As for the speed comparison of crowdsourcing translation and MT, an average of 68.6% believes that the speed in crowdsourcing is lower than MT (see Figure 6).
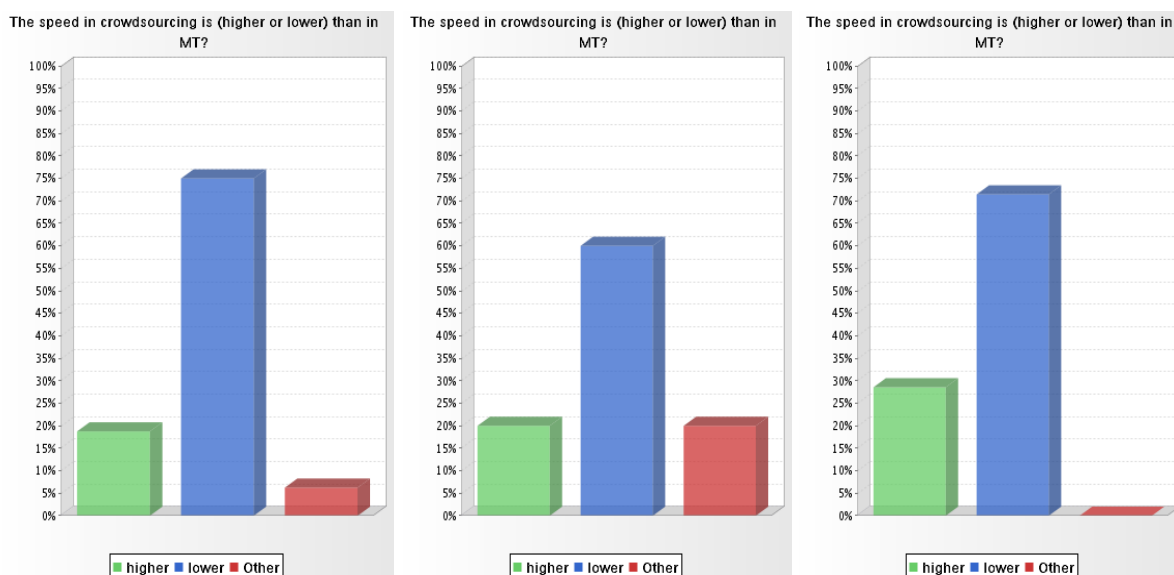
Figure 6. Responses on whether speed of crowdsourcing translation output is higher/lower quality than of MT (AsLiB – left, tc world – middle, students – right)

We deduce from the Figures 5 and 6 that the responses are similar and the results clear: the quality of crowdsourcing translation is higher than MT, while MT has higher speed than crowdsourcing translation. As concerns the volume, we see a differentiation of the responses (Figure 7). On the one hand, the predominant response from both conferences' respondents was that the volume in crowdsourcing is lower than MT, while on the other hand, the students believe that it is higher in crowdsourcing than MT:

Figure 7. Responses on whether the volume of crowdsourcing translation output is higher/lower than of MT (AsLiB – left, tc world – middle, students – right)

The respondents were also asked whether in their opinion crowdsourcing could bring down the profession of translation; on average 72% answered "no", followed by "maybe" and "yes". It should be pointed out that the students gave preference to "maybe" (49%) rather than "no" (38%), showing their slight fear of crowdsourcing 'intrusion'. Last but not least, the question whether crowdsourcing can compete with the current industry received the predominant response "no" by both conferences' participants and "yes" by the students; here the experience with industry and knowledge of industry trends should be considered.

6. Future Prospects

In this section we foresee some trends in the translation environment and make our suggestion combining automatic with human translation. The issues that we address in the next paragraphs are whether human translation should be neglected, thus preferring revision and postediting. Moreover, it will be discussed whether MT is as cheap solution for the company as it seems to be.

[30]'s statement that translation is an art is seriously fading away. One can argue that the creativity of the authors is sinking, particularly when the authors have to write based on style guidelines, such as these of controlled language (CL). Translators without any resources, tools, and technologies are a minority today. The work of translators becomes easier with use of tools and a restriction of vocabulary and grammar rules, but creativity and "translation art" is replaced by understandability and usability.

We now come to the future prospects of translation; we expect that the number of translators decreases and of proofreaders increases, as LSPs will send their content first to MT and TMs firstly to pretranslate the context and secondly to proofreaders for review. In crowdsourcing translation human translation is the key, so seeing translators only becoming proofreaders is automatically eliminated. MT depends on postediting and as [57] states "postediting will become a mainstream service, and word rates will increasingly be replaced by time-based or subscription and service-bundle pricing." The distinction between "postediting" and "revision" is worth mentioning. According to [58], postediting can be considered the correction and perfection of content already automatically translated by an MT system. The term "revision", which is although similar in some aspects, deals with the error correction of human produced draft translations. Postediting edits can involve, for example, correcting errors involving punctuation, inappropriate glosses, misconstructions of meaning, misspellings, mistakes in numerals, incorrect attachment and incorrect ambiguity resolutions amongst other things [58].

The relation between postediting and cost should be addressed here; it is not always that using MT and having the translated content postedited saves money to the company. According to [59], "if the MT system produces a perfectly valid translation, but the translator feels that it should have an alternative translation, we do not want the translator to perform any changes to the MT output, [because] (…) carrying out such unnecessary postediting effort reduces the benefits and potential cost savings of the MT system".

In the previous paragraphs we discussed revision and postediting which will come up in our following recommendation as a means to enhance quality.

In our attempt to make a recommendation for future workflows combining automatic and human translation, this is the paradigm we target at: low cost, high speed, and high quality (adapted from [47], see Figure 2).

High speed

Ensure all 3

Low price   High quality
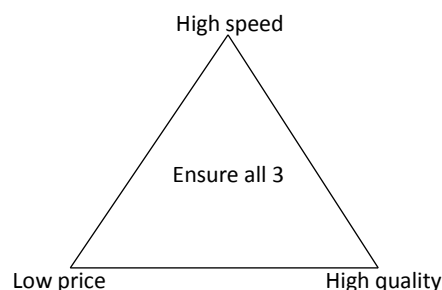
Diagram 8. The future localisation paradigm

We now suggest the ways in which we expect the ideal desired result: lowered price and enhanced speed and quality. As for speed, we mentioned previously that crowdsourcing platform is a cheap solution. In the beginning it certainly needs budget and good planning for its implementation, but in the long run, its maintenance is trivial, as it is upon the crowd/community to make the platform vivid. Also, MT and automatic translation tools obviously lower the cost of human translation. A wave of FOS systems being available today makes the access easy and breaks inequality and translators' discrimination barriers (those who can purchase a proprietary tool and those who do not).

Speed is enhanced by means of tools and also of a big number of registered crowd/community available and willing to translate content. Translators using FOS tools at a crowdsourcing platform is a good solution to accelerate the process.

High linguistic quality can be achieved through human involvement in translation and evaluation; despite the existence of quality evaluation metrics used for MT (BLEU, NIST, METEOR, TER, etc.), and crowdsourcing (Friend of a Friend (FOAF) algorithm)[17], both models can improve their output quality only through thorough human evaluation. This high linguistic quality can be tied with language consistency which is achieved through automatic translation tools. Often in crowdsourcing translation, translators do not use language resources, e.g. TMs, TDBs, glossaries, and generally relevant accumulating documentation; the result of that is inconsistent terminology and language. Here is where a marriage between humans and tools should be made; we plan initially to compare some translations (done with crowdsourcing translation and MT) with a gold-standard.

We bring the advantages of crowdsourcing translation (human translation, high linguistic quality) together with these of automatic tools (high speed, language consistency) and also low cost (characteristic of both) for our future translation workflow recommendation. We replace human-aided *machine translation* (HAMT) with human-aided *automatic* translation (HAAT), as nowadays we see a marriage between TM-MT-software localisation tools, resulting into so-called translation workstations or translation environment tools. In reality, when used for publishing purposes, professional users often combine MT and TMs. The simpler integration applies TMs first and then MT on "no match" segments as a default translation. A richer approach uses TM's and, more generally user feedback, to supply MT resources. We strongly believe that tools are not designed to work in isolation without human intervention. Thus we see future tools being automatic, but always with human involvement (translation and editing). The human translation can be sought at a crowdsourcing platform; thus our vision of HAAT is itself a combination of crowdsourcing translators with the characteristic of using and sharing references and databases. This sharing implies server-based TMs and TDBs which starts to become more popular the last years. After having the content translated by a human-aided tool, the content will go to the crowdsourcing community, which is the motivated, dedicated, and active part of the crowd. A controlled database with the best translators out of the crowd is necessary. This is feasible by asking for their CVs and their domain specialty at the log in screen. Professional translators can be involved in crowdsourcing by getting higher recognition and more suitable compensation on timely basis. The dedicated proofreader will use QA tools to facilitate their task and also rate the translations; having karma/stars as metadata next to the TM matches found will make a clear distinction between the highly and lowest rated translations, enhancing the translation output quality.

---

[17] In this paper we do not compare translation usings evaluation metrics (see future prospects – section 6).

With the workflow starting with crowdsourcing translation and HAAT, and followed by crowdsourcing proofreading, we combine high speed with high quality, resulting to a powerful, low cost, and flexible solution. Summarising, the steps of our proposed workflow are the following as they can be seen in Figure 9 as well:

1. **Users** enter a crowdsourcing/collaborative translation management platform and log in either as a translator or proofreader;
2. **PMs** take into account their credentials and give according recognition/compensation;
3. **Translators** pretranslate the content having an option to select FOS server-based TMs, MT systems, and software localisation tools, and consulting any relevant databases (previous TMs, etc.);
4. Complete the translation;
5. **Proofreaders** rate the translations at segment level;
6. Use **QA** tools;
7. **PMs** release the best-rated translations.
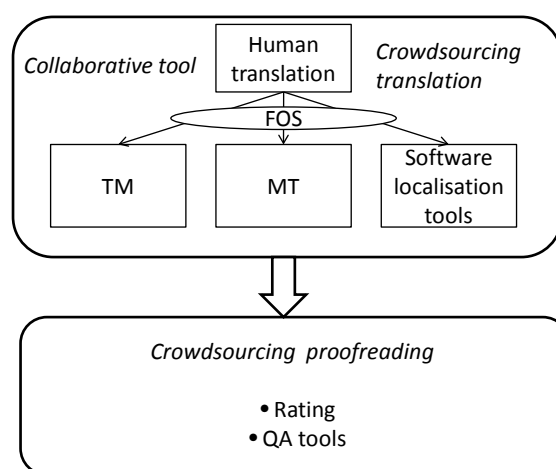
Figure 9. Crowdsourcing translation and proofreading

We should mention that *Asia Online*[18] uses SMT with postediting tools and integrates postediting with human collaboration environment. Among others, the postediting tools provide an environment to posteditors to identify the most common kinds of errors e.g. unknown words and make edit changes to the SMT system output to correct high frequency translation errors. These changes are directly fed back into the translation platform and thus Asia Online's Enterprise Translation Platform improves each time users provide feedback.

Also, a recently released collaborative translation tool called *Crowdin* is related with our recommendation combining crowdsourcing translation and TMs. Users upload their content to be translated, use TMs to pretranslate their content, and then edit and translate the remaining content. After that, users create a community and get them translate their content. Our difference with *Crowdin* is that people registering are actually the crowd/community who does the translation; it is not the vendor who requests translation.

As mentioned previously, also Microsoft announced the next version of Microsoft Translator technology which besides an update to Bing Translator, it consists of a customisable widget[19] that site owners can place on any webpage to make the page available in many languages, and also allows the owners to tailor the translations to their own content with the help of their community.

Crowdsourcing is certainly not a threat to MT; it is a productivity model with future and is here to stay, as real examples prove that; it depends on the companies' strategy and users' motivation and commitment to embrace its advantages and disadvantages, and bring them in symbiotic relationship to those of MT and generally automatic translation in order to achieve the desired result.

---

[18] http://www.asiaonline.net/Default.aspx, 05/02/11

[19]   http://blogs.msdn.com/b/translation/archive/2010/03/15/collaborative-translations-announcing-the-next-version-of-microsoft-translator-technology-v2-apis-and-widget.aspx, 05/02/11

7. Summary

In this paper we compared crowdsourcing translation with MT, two successful models. The fast time-to-market and simultaneous shipping of localisation products and services gives competitive advantage to the organisations and that is why they want a quick and flexible solution; also, the higher quality this solution has, the better.

We started with the definition of crowdsourcing by its pioneer, Howe, and after evaluating it, we provided our redefinition. Then more details and examples were furnished pertaining to crowdsourcing translation. Advantages and challenges of crowdsourcing related to quality, speed, and cost were also examined. The next section referred to MT and more precisely to a brief history and open MT (including open-source and MT services). Then we compared crowdsourcing with MT and examined whether the former can be a threat or form a symbiotic relationship with the latter. After describing our survey and its results, we arrived at some future prospects and our recommendations in the last section.

The early translation tools in 1950s aimed at FAHQT with neglecting the human intervention, and in 1980s more human-aided tools (HAMT, TMs) appeared. Today we see the new model of crowdsourcing to come and rock the boat changing the traditional defined translation workflow. Humans are certainly needed to resolve the problems irresolvable by MT, but also automatic tools are needed to accelerate the process and keep the terminology consistent.

Crowdsourcing translation (as a trend) and MT (as a technology) have their own benefits and pitfalls; however, they share many common characteristics, e.g. high volume, quick translation output, and low cost. This is the reason why we regard crowdsourcing as human MT (HMT). Organisations should be aware of both advantages and disadvantages of both models in order to make the best decision as for which model they will follow. Our proposed model is the combination of the two models: starting with automatic translation, not only MT, but also TMs, TDBs, software localisation tools, and generally translation environments, and finishing with proofreading by well selected and managed volunteer community.

Acknowledgment

Appendix

Crowdsourcing Questionnaire

1. You are a:
   - ☐ Tools provider
   - ☐ Content Publisher (Client)
   - ☐ Localisation Service Provider
   - ☐ Translator
   - ☐ Other: _____

2. In terms of localisation, would you translate something for free?
   - ☐ No
   - ☐ Yes (Circle some of the following reasons)

     - ➢ I saw an error in translation and wanted to fix
     - ➢ For a goodwill (which?)
     - ➢ To improve my job / career opportunities and make contacts
     - ➢ To gain a reputation in the localisation community
     - ➢ Other (add)

3. Are you aware of the concept of crowdsourcing?
   - ☐ Yes
   - ☐ No

4. If you are tools provider, do you have a crowdsourcing platform implemented?
   - ☐ Yes
   - ☐ No

☐ No, but I think of doing that in the future

5. What do you think about the quality of Crowdsourced translation?
☐ Very good, because it is human translation
☐ Good
☐ Average
☐ Not good
☐ Very bad

6. How can the quality of crowdsourcing be improved?
☐ Through terminology consistency (glossaries & terminology databases)
☐ Through administration security control
☐ Through manual proofreading
☐ Through checking that the translators are good enough
☐ Other: _____

7. Who do you think profits from crowdsourcing?
☐   The user
☐   The company
☐   The community

8. The quality in crowdsourcing is (higher/lower) than in MT?
☐   Higher
☐   Lower
☐   Other: _____

9. The speed in crowdsourcing is (higher or lower) than in MT?
☐   Higher
☐   Lower
☐   Other: _____

10. The volume in crowdsourcing is (higher/lower) than in MT?

☐   Higher
☐   Lower
☐   Other: _____

11. Provide any thoughts about the comparison of crowdsourcing to Machine Translation.

_____

12. Do you think crowdsourcing could bring down the profession of translation?
☐   Yes
☐   No
☐   Maybe

13. Is community based localisation (crowdsourcing) a good production model which can compete with the current industry?
☐   Yes
☐   No
☐   Maybe

14. Provide any thoughts, ideas, or recommendations about how you would motivate, build trust, compensate, and manage the crowd.

_____

References

[1] Zuckerman E. 'Jeff Howe on Crowdsourcing', http://www.ethanzuckerman.com/blog/2009/03/17/jeff-howe-on-crowdsourcing/ (2009, accessed February 2011).

[2] Howe, J. 'The Rise of crowdsourcing', *Wired*, Issue 14.06, http://www.wired.com/wired/archive/14.06/crowds.html (2006b, accessed February 2011).

[3] Howe J. 'Crowdsourcing: Tracking the Rise of the Amateur', http://crowdsourcing.typepad.com/cs/2006/06/customermade_th.html (2006c, accessed February 2011).

[4] Spigit. 'Four Models for Competitive Crowdsourcing', http://www.spigit.com/2010/09/four-models-for-competitive-crowdsourcing/ (2010, accessed March 2010).

[5] Brabham D.C. Crowdslapping the Government: First Amendment Protections for the Crowd in Government crowdsourcing Ventures. *Annual meeting of the Association for Education in Journalism and Mass Communication,* Chicago, Illinois, 2008, pp.1-32.

[6] Sample Ward A. 'Crowd vs Community', http://amysampleward.org/2010/03/14/crowdsourcing-community-vs-crowd/ (2010, accessed February 2011).

[7] Haythornthwaite C. Crowds and communities: Light and heavyweight models of peer production. *Proceedings of the 42ⁿᵈ Annual Hawaii International Conference on System Sciences (HICSS)*, Waikoloa, HI, USA, 2009, pp.1-10.

[8] Howe, J. 'Crowdsourcing', http://www.youtube.com/watch?v=F0-UtNg3ots (2008, accessed June 2010).

[9] Meier P.P. 'Haiti and the Power of Crowdsourcing', http://irevolution.net/2010/01/26/haiti-power-of-crowdsourcing/ (2010, accessed February 2011).

[10] Callison-Burch C and Dredze M. Creating Speech and Language Data with Amazon's Mechanical Turk. *Proceedings NAACL Workshop Creating Speech and Language Data with Amazon's Mechanical Turk*, 2010.

[11] DiNucci D. *Fragmented Future*. 1999, Print 53 (4): 32.

[12] Albors J, Ramos J and Hervas J. New learning network paradigms: communities of objectives, crowdsourcing, wikis and open source. *International Journal of Information Management,* 28(3), 2008, pp.194-202.

[13] Google In Your Language: https://www.google.com/accounts/NewServiceAccount?service=transconsole&continue=http%3A%2F%2Fwww.google.com%2Ftransconsole&followup=http%3A%2F%2Fwww.google.com%2Ftransconsole (accessed June 2010).

[14] Sawers P. Facebook's un-rebellion. *Multilingual April/May 2009*, 2009, 62.

[15] Petition: Professional Translators Against Crowdsourcing and Other Unethical Business Practices http://www.petitiononline.com/TEBP3/petition.html, (accessed June 2010).

[16] TED Open Translation Project. http://www.ted.com/OpenTranslationProject

[17] Worldwide Lexicon. http://www.worldwidelexicon.org/ (accessed February 2011).

[18] CoSyne. http://www.cosyne.eu/index.php/Main_Page (accessed February 2011).

[19] MediaWiki. http://www.mediawiki.org/wiki/Main_Page (accessed February 2011).

[20] Raymond E. (1998). *Goodbye, free software; hello, open source.*

[21] Cronin M. The Translation Crowd. *revista tradumàtica,* 08, Localitzaciò i web, December 2010.

[22] Wolf T, Nguyen T and Damian D. Does distance still matter?. *Software Process: Improvement and Practice*, 13 (6), 2008, pp. 493-510.

[23] Kittur A. and Kraut R.E. Harnessing the Wisdom of Crowds in Wikipedia: Quality Through Coordination. *Proceedings of the ACM 2008 conference on Computer supported cooperative work*, New York, NY, USA, 2008, pp. 37-46.

[24] Ipeirotis P. (2010), Mechanical Turk: Now with 40.92% spam, A Computer Scientist in a Business School http://behind-the-enemy-lines.blogspot.com/2010/12/mechanical-turk-now-with-4092-spam.html (2010, accessed February 2011).

[25] Huberman B. Crowdsourcing and Attention. *Computer*, 41(11), 2008, pp.103-105.

[26] Moore T.D. and Serva M.A. Understanding member motivation for contributing to different types of virtual communities: a proposed framework. *Proceedings of the 2007 ACM SIGMIS CPR conference on computer personnel research: The global information technology workforce*, St. Louis, Missouri, USA, 2007, pp.153-158.

[27] Kuznetsov S. Motivations of contributors to Wikipedia. *SIGCAS Comput. Soc.*, 36(2), 2006, 1.

[28] Nielsen J. Participation Inequality: Encouraging More Users to Contribute, http://www.useit.com/alertbox/participation_inequality.html, (2006, accessed February 2011).

[29] Kostakos V. Is the Crowd's Wisdom Biased? A Quantitative Analysis of Three Online Communities. *Proceedings of the 2009 International Conference on Computational Science and Engineering.* Volume 04. IEEE Computer Society, 2009, pp. 251-255.

[30] Hoffmann L. Crowd control. *Communications of the ACM*, 52(3), 2009, pp. 16-17.

[31] Hutchins W.J. and Lovtskii, E. Petr Petrovich Troyanskii (1894-1950): a forgotten pioneer of mechanical translation. *Machine Translation,* 15 (3), 2000, pp. 187-221.

[32] Hutchins W.J. *Machine translation: past, present, future.* Chichester: Ellis Horwood. New York: Halsted Press,1986.

[33] Holmström J.E. *Report on interlingual scientific and technical dictionaries*. Paris: Unesco, 1951.

[34] SYSTRAN Enterprise Server 7. http://www.systran.co.uk/translation-products/server/systran-enterprise-server and http://www.systran.co.uk/translation-products/server/systran-enterprise-server/language-pairs (accessed February 2011).

[35] Pierce J.R., Carroll J.B., Hamp E.P, Hays D.G, Hockett C.F, Dettinger A.G. and Perlis A. *Language and Machines. Computers in Translation and Linguistics*, Report by the Automatic Language Processing Advisory Committee (ALPAC), Division of Behavioral Sciences, National Academy of Sciences, National Research Council, Washington, D.C.: National Academy of Sciences, National Research Council, 1966, Publication 1416.

[36] Kay M. Machine Translation. *American Journal of Computational Linguistics*, 8 (2), 1982, pp. 74-78.

[37] Kay M. The Proper Place of Men and Machines in Language Translation. *Machine Translation,* 12, 1997, pp. 3-23.

[38] Brown P, Cocke J, Pietra D, Pietra V, Jelinek F, Mercer R, Roossin P. A Statistical Approach to Language Translation. *12th COLING* 1988, Budapest, Hungary*,* 1988, pp.71-76.

[39] van der Meer J. Fully Automatic Useful Translation. *Proceedings of EAMT 11th Annual Conference*, 2009, p. 5.

[40] Hutchins W.J. 'Compendium of Translation Software', http://www.hutchinsweb.me.uk/Compendium.htm (2009, accessed June 2010).

[41] Forcada M. Open source machine translation: an opportunity for minor languages. *Proceedings of the 5th SALTMIL Workshop on Minority Languages: "Strategies for developing machine translation for minority languages", 5th International Conference on Language Resources and Evaluation (LREC-2006)*, Genoa, Italy, 2006, pp. 1-6.

[42] Papineni K, Roukos S, Ward T and Zhu W.J. BLEU: A method for automatic evaluation of machine translation. *40th ACL 2001*, Philadelphia, Pennsylvania, 2001, pp. 311-318.

[43] Och F.J and Ney H. Improved Statistical Alignment Models. *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, Hongkong, China, 2000, pp. 440-447.

[44] Dandapat S, Forcada M.L, Groves D, Penkale S, Tinsley J and Way A. OpenMaTrEx: A Free/Open-Source Marker-Driven Example-Based Machine Translation System, *Proceedings of the 7th International Conference on Natural Language Processing (IceTAL),* Reykjavík, Iceland, 2010.

[45] Joscelyne A. Open MT: ready for business? http://www.translationautomation.com/technology/open-mt-ready-for-business.html (2009, accessed February 2011).

[46] Koehn P. A Web-Based Interactive Computer Aided Translation Tool. *Proceedings of the ACL Interactive Poster and Demonstration Sessions*, Singapore, 2009.

[47] Thicke L. Optimized MT for Higher Translation Quality. *Multilingual October/November 2009*, 9-11.

[48] DePalma D. A. and Kelly N. The Business Case for Machine Translation. http://www.mt-archive.info/MTS-2009-DePalma-ppt.pdf (2009, accessed February 2011).

[49] Flournoy R. and Duran C. Machine translation and document localization at Adobe: From pilot to production. *MT Summit XII: proceedings of the twelfth Machine Translation Summit*, Ottawa, Ontario, Canada, 2009.

[50] van der Meer J. (2009). The Profit of Sharing. Let a Thousand MT systems bloom. *TAUS User Conference 2009*, Portland, USA.

[51] Language Technology Centre (LTC). 'Study Report to the Directorate General for Translation of the European Commission', http://ec.europa.eu/education/languages/news/news3725/report_en.pdf (2009, accessed February 2011).

[52] Cooper C. 'Is Translation Crowdsourcing Unethical? ', http://colincooper.net/?p=31 (2009, accessed February 2011).

[53] Anastasiou D. Crowdsourcing and Machine Translation. Poster in *Proceedings of the Localisation Research Centre LRC XIV, Localisation in the Cloud*, Limerick, Ireland, 2009.

[54] Muegge U. Fully automatic high quality machine translation of restricted text: A case study. *Proceedings of the 28th International Conference on Translating and the Computer*, London, UK, 2006.

[55] Braster B. Controlled language in technical writing. *Multilingual January/February 2009*, 2009, pp. 50-52.

[56] Di Maio A. 'What's the Real Cost of Crowdsourcing?'. Gartner blog, http://blogs.gartner.com/andrea_dimaio/2009/06/04/whats-the-real-cost-of-crowdsourcing/ (2009, accessed February 2011).

[57] Allen J. Post-editing. *Computers and Translation: A Translators Guide*, In: Somers H. (Ed.), Benjamins Translation Library, 35. Amsterdam: John Benjamins, 2003, pp. 297-317.

[58] McElhaney T. and Vasconcellos M. The translator and the postediting experience. In: M. Vasconcellos (Ed.) *Technology as Translation Strategy* (American Translators Association Scholarly Monograph Series, vol. II), Binghamton, New York: State University of New York, l988.

[59] Groves D. Bringing Humans into the Loop: Localization with Machine Translation at Traslan. *Proceedings of the American Association for Machine Translation (AMTA)*, 2008, pp.11-22.

Corresponding author:
Dimitra Anastasiou, CARTESIUM 1.13, Enrique-Schmidt Str. 5, 28359 Bremen, anastasiou@uni-bremen.de