Managing Social Translation: Online Tools for Translators' Communities

Anas Tawileh

VP Engineering, Meedan.net

1   Introduction

The rapid technological development over the past few decades has significantly changed many aspects of human lives. The advent of digital technologies transformed the ways in which people communicate, conduct business and entertain. The digital revolution is offering unprecedented opportunities for businesses and organizations to create and access new markets and maximize their productivity and profitability. For the professional translation community, which is a profession based primarily on knowledge and experience, with little dependence on capital investments and tangible assets, these developments provide unprecedented opportunities to increase the efficiency and productivity of translation work, and open new markets or improve access to existing ones across geographic borders. Moreover, the digital lifestyle of more than a billion citizens of the world is generating massive amounts of information stored in digital format and made available on the web. According to a recent study by consultancy firm International Data Corporation (IDC), the world's digital output currently stands at 8,000,000 petabytes and may surpass 1.2 zettabytes [1] (a zettabyte is equal one million petabytes). This creates substantial demand for translators and translation services, but also poses challenges to the traditional approaches to translation management and conduct, and the way translators have organized in the past.

Traditional translation workflows are usually more focused on translating large chunks of static content, and are therefore severely limited in their ability to translate online dynamic content. This is primarily caused by several characteristics of the emerging processes for content creation and the nature of their communities. The Internet enables instantaneous communication that transcends geographical boundaries. Larger, more diverse communities can now leverage this new medium to organize around topic of mutual interest. Empowered by the democratic means of content and knowledge production on the social web, these communities are collaboratively producing massive amounts of content. Probably the most notable example of the products of these diverse, large-scale and highly collaborative production processes is

Wikipedia, the free and open online encyclopedia written by hundreds of thousands of volunteers distributed all over the globe. As of the date of the paper, Wikipedia includes around 3.4 million articles, totaling more than one billion words [2]. In another example, users of the popular photo sharing website Flickr [3] are uploading 6,000 new photos on average.

But the massive scale of peer content production on the social web is only one aspect of the social web. The nature of the web and digital media facilitates continuous change and modification of content. This is reflected in more dynamic and rapidly changing online content, in stark contrast with the stability and rather lengthy update cycles for content published in physical mediums, such as books and videotapes. In an article by Ethan Zuckerman, he argued that we are "likely to miss the diversity and nuance of the user-generated web unless we found better ways to translate the variety of languages we're seeing online" [4].

To address this issue, several projects were initiated to develop innovative approaches to translation that harness the power of the emerging technological developments like the social web, mobile connectivity and machine translation, to cater for the requirements of modern translator communities. The resulting approach, termed "social translation:" blends the speed and cost-effectiveness of machine translation with the accuracy and quality of human translation to provide translators with significant efficiency and productivity gains, and enable them to respond to the changing dynamics in today's translation marketplace. The next sections explore these concepts in greater detail.

2   Social Translation

Social translation is an innovative approach that responds to the rapidly changing trends in content production and dissemination by transforming the translation process to become more open, inclusive and collaborative. The new approach facilitates collaboration among large numbers of translators, sometimes called "translator communities", so that they can work together on translating large, dynamics and continually changing content. Its ultimate goal is to increase productivity and efficiency of the translator's work by empowering distributed translator communities.

Several components provide the functionality required to enable social translation. These include machine translation engines, reputation systems, accounting and micro-payment, translation rating mechanisms, glossary tools and linguistic corpus and translation memory.

While much has been said about developments in machine translation technologies, the reality remains that these translations tend to be of lower quality that human translations. Frequently, content consumers are complaining that machine translation has not yet reached sufficient maturity levels to produce meaningful content. However, machine translation can still provide a viable starting point for translators to work from and improve. This is particularly valuable in specific fields such as news reporting, where machine translation can achieve feasible learning.

For this purpose, social translation builds on initial translation drafts generated by machine translation engines, and empowers translators' communities with the ability to vet the translation and make any necessary corrections and amendments. After an initial round of machine translation of the source content, the translated materials are distributed redundantly in small chunks to a network of distributed translators for correction or revision. The updated segments are then collected and collated to produce the first human-revised version of the translation.

2.1   Task Allocation and Management

To manage the social translation process, an effective task allocation and management mechanism must be available. In the Hybrid Distributed Natural Language Translation (HDNLT) approach developed by Meedan, when a piece of content is designated as needing translation, a flag will be added to this piece that will trigger the task allocation process. This step can also include specific service level requirements, such as the translator's specialization and reputation, as well as the requested turnaround time. Part of this task management workflow includes the ability to assign document security parameters and document translation urgency. At this stage, the type of service can also be specified: human only, machine only or hybrid. Based on these specifications, the system will either perform an initial machine translation of the submitted content or bypass this stage. The system then creates a translation request and matches the task with properly qualified and available translators, and sends the request to the highest ranking translator. If the translator did not respond within a specific time to live value, the request is rerouted to the next available translator until a translator accepts the task allocation.

The system provides an intuitive, easy to use interface to facilitate the translator's work, including access to the translation memory and other glossary management tools. Upon completion of the translation task, the translator submits the translated materials back to task management system, which will in turn ensure that the content is properly formatted, and send it

back to the originating content management system. This process minimizes manual interaction during task execution, and can be easily managed through a high level management interface

2.2   Quality and Reputation Management

Opening up the translation process to a community of translators poses an important challenge relating to the quality of the contributed translations. How can the system ensure that translations submitted by the community are of acceptable quality? And how does it deal with those who consistently produce poor translations?

Social translation tools and platforms address this by designing a portable translation reputation management system for translators, coupled with a rating mechanism for translated content. Readers can use the rating mechanism to give feedback on the quality of the translated materials. This feedback is aggregated to calculate the reputation score for each translator. Future routing decisions of translation requests will be made based on the reputation scores of translators in the system. This mechanism is vital to provide an incentives structure that promotes quality within the system, and mitigates the potential for gaming or malicious intent. Translators who produce higher quality output will be more likely to get more work, and possibly command higher compensation for their work. The reputation score decays over time, which serves as a disincentive for translators not to rely on their previous performance, and consistently deliver high quality output. To maximize the utility of this structure, the translator's reputation is designed as a portable attribute to enable its use across different systems. We perceive the development of a standard translator reputation management protocol to facilitate the interoperability of reputation scores between different systems and platforms. For example, a translator may include her portable reputation score in her LinkedIn profile to communicate the quality of her work.

2.3   Tracking and Invoicing

As the volume of content translated on the platform increases, the administrative load associated with the process should be reduced. This will help the translators' community focus on their core task, and will eliminate the burden on the end user, as most of the required support transactions will be automated. The tracking and invoicing component of social translation platforms should automatically calculate the size of translated content, match that number with the translator's rate, and generate automated invoices for the work performed. The user can then establish a review process for invoices, and can disburse payments through the payment management system. A

configurable notification system can also be integrated to enable users to receive notifications when new invoices are submitted.

2.4　Payment Management

The purpose of this component is to ensure that payments for completed translation tasks are issued in a correct and timely manner. Social translation platforms can integrate with popular online micropayment services, such as PayPal and MoneyBookers, so that approved payments can be issued to their respective translators. This will facilitate swift and timely payment, and will reduce the administrative overhead for the user. Both translators and users can review detailed histories of payments, and generate relevant reports.

2.5　Glossary and translation memory management

This component intends primarily to improve the translators' efficiency and productivity by offering them instant access to translation repositories. Translators can utilize these tools to retrieve exact or similar translations for specific segments from defined translation memories, or to inquire about the specific meaning of terms in the materials they are translating.

2.6　Application Programming Interfaces (APIs)

The Application Programming Interfaces (APIs) are well defined interfaces that can be utilized to access the functions and features of the social translation platform. For example, users can use an API to trigger the translation process. When a document is deemed to need translation, a flag can be added to this document in the user's content management system which will automatically submit the document to the translation marketplace, along with any service level requirements. When the document is translated, it will be automatically published back on the user's content management system through another API. By utilizing APIs, modifications of existing systems will be minimized, and developers do not have to spend excessive time learning the internals of different applications. When a feature or function is required from the social translation platform, the associated API will be used to request that feature.

3　Social Translation Projects

Several initiatives and projects have been launched to provide platforms and tools for social translation on the web. The objectives of these projects vary widely from general purpose platforms to initiatives focused on specific types of content, such as education, news or science. Another dimension across which these projects differ is the languages they enable. While some projects

are focused on specific language pairs, others cater for higher numbers of languages. The next sections present two open source projects for social translation.

3.1   Meedan

Meedan is a non-profit based in San Francisco, California that specializes in advanced social translation technologies and focuses on the English and Arabic languages. Meedan leverages its patented Hybrid Distributed Natural Language Translation (HDNLT) technology to build collaborative translation platforms to enable translators' communities to move content across the language barrier. Meedan's technologies have been implemented by several projects to enable cross-language sharing of news and media (news.meedan.net) [5], empower cross-language inter-faith dialogue among prominent faith thought leaders [6], and support translation of education content and connect classrooms [7].

All of the translation technologies developed by Meedan are released under an open source software license to maximize their value and facilitate further development. The linguistic data produced by its translators' communities is also made available under open data licenses so that it can be leveraged by other projects and communities.

3.2   Pootle

Pootle is another project that provides tools to empower social translation communities. It offers a "user-friendly web portal that makes the translation process so much simpler. It allows online translation, work assignment, gives statistics and allows easy volunteer contribution." [8]

Pootle can either be used as an Internet server, such as those implemented for the localization of OpenOffice [9], the open source office productivity suite, and the Mozilla localization project [10]. It can also be installed as an enterprise server for specific organizations or communities. As of today, Pootle is being used by more than 200 organizations and projects around the world.

4   Trends Social Translation

While several initiatives are currently underway to explore the potential and applicability of social translation in addressing the challenges of translating the massive amount of dynamic content created on the web, and pretty much like the social web, social translation is an emerging phenomenon that is consciously changing to reflect developments on several fronts. As we move

forward, the maturity and feasibility of machine translation is expected to increase, which will fundamentally change the role of the translator. We will also see increased utilization of translation memories and online glossaries as more content is created through social translation platforms and computer aided translation tools (CAT) such as Google Translator Toolkit. This will have major implications on improving the productivity of the translator's work, and increasing the cost effectiveness of translation activities.

More social translation projects and platforms will be built around different language communities and specific subject area specialization, which will lead to increased pressure for the development and adoption of standardized quality assessment mechanisms and reputation scores. This standardization will facilitate the interconnectivity of different social translation platforms, and give users of social translation services more choice.

Additionally, the increasingly web-based approach to social translation will enable the use of new computing platforms by translators. For example, translators may soon be able to use social translation clients for mobile phones or tablet computers to process their translation tasks and connect with their online communities. Meedan is currently experimenting with the developments of SMS-based social translation tools.

While these trends and developments will undoubtedly affect the translation profession as it is known today, they also offer new exciting opportunities and prospects for translators willing to embrace change and adopt new approaches to the way they work. Currently, for every English speaker on the web, there are two others who do not speak English [11]. As we strive to add the next billion users, most of which do not speak English, the role of the translator, particularly those involved in social translation communities, will be more vital than ever.

5   References

[1]   New Statesman (2010), Zettabyte is new unit of measuring digital output, http://www.newstatesman.com/digital/2010/05/digital-output-word-zettabyte, , accessed 25 October 2010.

[2]   http://en.wikipedia.org/wiki/Wikipedia:Size_comparisons, accessed 25 October 2010.

[3]   http://www.flickr.com, accessed 25 October 2010.

[4]   Ethan Zuckerman (2008), New York Times on Social Translation, http://www.worldchanging.com/archives/009893.html, , accessed 25 October 2010.

[5]   http://news.meedan.net , accessed 25 October 2010.

[6]   Cambridge Interfaith Programme, http://www.interfaith.cam.ac.uk/en/cap/abraham-online, accessed 25 October 2010.

[7]   QFI (2010), Qatar Foundation International announces new program forging relations between Arab and American youth, http://www.qfi.org/archives/1647, accessed 25 October 2010.

[8]   Pootle, http://translate.sourceforge.net/wiki/pootle/index?redirect=1, accessed 25 October 2010.

[9]   Pootle for OpenOffice.org, http://pootle.sunvirtuallab.com/, accessed 25 October 2010.

[10] Mozilla Verbatim, https://localize.mozilla.org/, accessed 25 October 2010.

[11] ByteLevel (2005), English-Speaking Web Users a Minority on the Internet, http://www.bytelevel.com/news/web_languages_nov3_2005.html, accessed 25 October 2010.