AUTOMATED TRANSLATION OF A LITERARY WORK: A PILOT STUDY

Author 1
XYZ Company
111 Anywhere Street
Mytown, NY 10000, USA
author1@xyz.org

Author 2
ABC University
900 Main Street
Ourcity, PQ, Canada A1A 1T2
author2@abc.ca

Abstract

Current machine translation (MT) techniques are continuously improving. In specific areas, post-editing (PE) allows to obtain high-quality translations relatively quickly. But is such a pipeline (MT+PE) usable to translate a literary work (fiction, short story) ? This paper tries to bring a preliminary answer to this question. A short story by American writer Richard Powers, still not available in French, is automatically translated and post-edited and then revised by non-professional translators. Our post-editing platform allows to read and edit the short story suggesting (for the future) a community of readers-editors that continuously improve the translations of their favorite author. In addition to presenting experimental evaluation results of the MT+PE pipeline (MT system used, automatic evaluation), we also discuss the quality of the translation output from the perspective of a panel of readers (who read the translated short story in French, and answered to a survey afterwards). Finally, some remarks of the official french translator of R. Powers, requested on this occasion, are given at the end of this article.

Index Terms— machine translation, MT, literature, fiction, post-edition

1. INTRODUCTION

The task of post-editing consists of editing some text (generally produced by a machine, such as a machine translation, optical character recognition, or automatic transcription system) in order to improve it. When using machine translation in the field of document translation, the following process is generally used: the MT system produces raw translations, which are manually post-edited by trained professional translators (posteditors) who correct translation errors. Several studies have shown the benefits of the combined use of machine translation and manual post-editing (MT+PE) for a document translation task. For example, (Garcia, 2011) showed that even though post-editing raw translations does not always lead to significant increases in productivity, this process can result in higher quality translations (when compared to translating from scratch)[1]. Autodesk also carried out an experiment to test whether the use of MT would improve the productivity of translators. Results from that experiment [2] show that post-editing machine translation output significantly increases productivity when compared to translating a document "from scratch". This result was true regardless of the language pair, the experience level of the translator, and the translator's stated preference for post-editing or translating from scratch.

These results from academia ((Garcia, 2011)) and industry (Autodesk) regarding translation in specialized areas lead us to ask the following questions:

- What would be the value of such a process (MT + PE) applied to the translation of a literary work?
- How long does it take to translate a literary document of ten thousand words?
- Is the resulting translation acceptable to readers?
- What would the official translator (of the considered author) think of it?

[1] the work of (Garcia, 2011) is however subject to controversy because the manual translation without post-editing appears to have been done without authorizing the use of digital assistance to help the translator (not even an electronic dictionary)

[2] amta2012.amtaweb.org/AMTA2012Files/html/5/5_paper.pdf

- Is "low cost" translation produced by communities of fans (as is the case for TV series) possible for novels or short stories?

This article attempts to provide initial answers (albiet very preliminary) to these questions. In addition to this preliminary study, this work will at the least provide a new translation of a short story or rather an essay (*The Book of Me* by Richard Powers).

We want to mention that this work is a tranlsated and extended version of a paper published in French in a local NLP conference last year[3].

This paper is organized as follows : next section presents our general methodology, the choice of literary work considered and the two systems used, MT and PE (both are "inhouse" systems). The experimental data collected (and made available to the community on *github*) are then described in Section 3. In part 4, we try to go beyond the MT domain-adaptation task by requesting the point of view of one (small) panel of readers and the opinion of the official translator of R. Powers. Finally, Section 5 concludes this preliminary work.

2. AUTOMATED TRANSLATION OF A LITERARY WORK

2.1. Preliminary remarks

During our literature review on assisted translation of literary documents, we did not find many works except (Voigt and Jurafsky, 2012) published in the *ACL workshop on Computational Linguistics for Literature*[4] (this workshop deals with the use of NLP in the literary field since 2012). In (Voigt and Jurafsky, 2012), the authors examine how referential cohesion is expressed in literary and non-literary texts and how this cohesion affects translation (experiments on Chinese literature and news). The present paper, however, tries to investigate if computer-assisted translation of a complete (and initially un-translated) short story, is reasonnably possible or not.

Also, we should now define what constitutes, for us, a literary text. For our purposes, we include in this category (our definition is undoubtedly too restrictive) all fiction or autobiographical writing in the form of novels, short stories or essays. In these texts, the author expresses his vision of the world of his time and life in general while using literary devices and a writing technique (form) that allows him to create effects using the language and to express meanings (explicit or implied).

2.2. Choice of the short story

The choice of literary work has been guided by the fact that

(a) we had a contact with the French translator of the American author Richard Powers [5] (author of the novel *The Echo Maker* which won the National Book Award and was a finalist for the Pulitzer Prize)

(b) R. Powers often explores, in his writings, the effects of modern science and technology, and in some ways his writings contain commonalities with scientific and technical texts (which may somewhat reduce the gap between translation of scientific and literary texts).

Via his French translator (J-Y Pellegrin), R. Powers was informed by e-mail of our approach, and he gave his consent as well as his feeling on this project (R. Powers: *"../... this automated translation project sounds fascinating. I know that the field has taken a big jump in recent years, but each jump just furthers the sense of how overwhelming the basic task is. I would be delighted to let him do a text of mine. Such figurative writing would be a good test, to be sure. .../... The Book of Me would be fine, too. "*).

We then chose a text not yet translated into French, entitled *The Book of Me*, originally published in the pages of GQ magazine [6]. It is a narrative, written in the first person, where the author is the main character of the story. He tells of the year 2008 during which he became the ninth person in the world to see his fully sequenced genome. Although the topic is genetics and in spite of the simple, clinical style used by the author, *The Book of Me* is truly a work of literature in which the author, who teaches narrative technique at the university, never puts aside his poetic ambition, his humour and his fascination for the impact of science and technology on the society.

[3] https://aclweb.org/anthology//F/F14/

[4] sites.google.com/site/clfl2014a/

[5] en.wikipedia.org/wiki/Richard_Powers

[6] www.gq.com/news-politics/big-issues/200810/richard-powers-genome-sequence

2.3. General methodology

The translation of texts using the MT+PE (machine translation and manual post-editing) workflow has previously been evaluated by us on journalistic data. In 2012, 12,000 post-edited segments (equivalent to a book of about 500 pages) were collected by (XXX, 2012). This corpus, collected through crowdsourcing and available online [7] , is one of the largest existing corpus of post-edited machine translation applied to freely available data. It is, for example, three times larger than that collected by (Specia et al., 2010), which is one well known benchmark in the field. This same methodology was applied to our chosen literary work but (significant difference) there was only one post-editor (no crowdsourcing) and post-edited text was subsequently revised.

Specifically, the corpus was divided into three equal parts. A translation/post-edition/adaptation loop was applied on the three blocks of text according to the following process:

- The post-edited first third was used to adapt the English-French MT system. Given the small amount of post-edited data, the adaptation only consists in adapting the weights of the log-linear SMT model (by using the corrected first third as a development corpus). A similar method is suggested by (Pecina et al., 2012) for domain adaptation with a limited quantity of data (we are aware that other more advanced domain adaptation techniques could have been used but this was not the central theme of our contribution),

- Then, the second third of the text was translated with the adapted MT system, then the results were post-edited and a second adapted MT system was obtained starting from the new data. This second system was used to translate the third and last part of the text,

- Once the post-edition (PE) was completed, the final text was revised: first by the post-editor and then by another reviewer (not a professional, a French-speaking person with a good knowledge of the English language)

- Times of post-editing and revision were measured.

2.4. MT system used

Our system is trained using the data provided in the IWSLT evaluation campaign and optimized for the English-French MT task of the 2012 campaign. The following corpora are used to build the translation model (cumulative total of about 25M sentences): *news-c, europarl, un, ted, eu-const, dgt-tm, pct* and 5M sentences extracted from the *gigaword* corpus . The French part of the same corpus is used to train the language model, with the addition of the *news-shuffle* corpus provided as part of the WMT 2012 campaign. Our system is a phrase-based system and uses the *Moses* (Hoang et al., 2007) toolkit. Three translation models learned from different corpora (*ted ; news-c+europarl+un+eu-const+dgt-tm+pct ; gigaword5M*) are used. A 5-gram language model with modified Kneser-Ney smoothing is learned separately for each corpus using the SRILM toolkit (Stolcke, 2002) ; these models are then interpolated by optimizing perplexity on the IWSLT dev2010 corpus. The system obtains BLEU scores of 36.88 and 37.58 on the IWSLT tst2011 and test2012 corpora, respectively (BLEU evaluated with case and punctuation). More details on this system (which ranked respectably in IWSLT 2012) can be found in (AAA, 2012). It is clear that this system is not optimal for translating literary texts, and it would be desirable for future work to at least collect literary texts in French to adapt the target language model (or even consider having access to other works and translations of the same author).

2.5. Post-editing

We use our inhouse post-editing interface initially proposed by (SSSS, 2008). It has been used for many projects: translation of the EOLLS encyclopedia , multilingual access to dozens of websites (80 demonstrations, 4 industrial contracts, 10 target languages, 820k post-edited segments) - see more details in (BB, 2013).

Figure 1 shows the post-editing interface in advanced mode that allows, for each source segment, multiple automatic translations (from Google, Moses[8] , etc..) to be loaded and corrected . The

[7] link anonymized

[8] for this experiment, the output of our Moses system was

post-editing times are measured, and the history of corrections is stored in a database. This tool is the backbone that gave rise to the iMAG concept (*interactive Multilingual Access Gateway*) proposed by our team which allows browsing a site or subsite in several access languages, with incremental improvement and quality control of the translations (see link at footnote [9] for more details).

3. EXPERIMENTAL DATA

3.1. Corpus and post-editing statistics

The data, made up of 545 segments and 10731 words was divided into three equal blocks. Table 1 summarizes the number of source and target (MT or PE[10]) words in the data. Not surprisingly, a ratio greater than 1.2 is observed between French target (MT) and English source words. However, this ratio tends to decrease after post-editing of the French output.

3.2. Performance of the MT system

Table 2 summarizes the MT performance (measured with BLEU[11]) calculated on the full corpus with the systems resulting from each iteration. Post-editing time required for each block is also shown. The BLEU scores, which are directly comparable (because evaluated on the full corpus), show no real improvement of the system. It therefore appears that adaptation of weights alone (which resulted in improvements in (Pecina et al., 2012)) is ineffective in our case. However, post-editing time decreases slightly with each iteration (but again, the differences are small and it is unclear whether the decrease in post-editing time is due to the adaptation of the MT system or to increasing productivity as the post-editor adapts to the task). In the end, the total PE time is estimated at about 15 hours.

displayed in priority

[9] footnote anonymized

[10] The post-editing used here is obtained after each iteration of the process; the last stage of revision is thus not taken into account at this stage.

[11] The results reported in our French paper are slightly different since they were measured with a different pre-processing on the data - the conclusion is similar anyway

3.3. Analyzing the revised text

Reading the translated work at this stage (after PE) is unsatisfactory. Indeed, the post-editing is done "segment by segment" without the context of the full corpus. This results in a very embarrassing lack of homogeneity for a literary text. For this reason, two revisions of the translated text are also conducted: one by the post-editor himself (4 hours) and one by the author of this article (6 hours). The final version of the translated work (which has been obtained after 15+4+6=25 hours of work) provides the basis for more qualitative assessments which are presented in the next section. The difference between the rough post-edited version (15 hours of work) and the revised version (25 hours of work) is analyzed in Table 3. It is interesting to see that while the revision takes 40% of the total time, the revised text remains very similar to the post-edited text (BLEU score between PE and REV is 79,92 which is very high). So post-edition and revision are very different tasks and this is illustrated by the numbers of Table 3. More qualitative analysis of the revised text and its comparison with post-edited text is part of future work (and any reader interested in doing so can download our data set on *github*).

4. GOING BEYOND CLASSICAL MT DOMAIN ADAPTATION

4.1. The views of readers on the post-edited translation

Nine French readers agreed to read translated work and answered a questionnaire, available on *fluidsurveys.com*[12]. A pdf version of the test results and a spreadsheet file containing the results of the survey are also made available on *github*. After three questions to better understand the profile of the player (*How old are you? Do you frequently read? If yes, what is your best reading genre?*), the first portion (5 questions) asks readers about readability and quality of the translated literary text (*What do you think about text readability? Is the text easy to understand? Does the language sound natural? Do you think sentences are correct (syntactically)? Did you*

[12] https://fluidsurveys.com/surveys/manuela-cristina/un-livre-sur-moi-qualite-de-la-traduction/?TEST_DATA=

Fig. 1. Post-editing interface in advanced mode

Table 1. Source corpus, translated target, and corrected target

Iteration (no. seg)	English (no. words)	French MT (no. words)	French PE (no. words)
It.1 (184)	3593	4295	4013
It.2 (185)	3729	4593	4202
It.3 (176)	3409	4429	3912
Total (545)	**10731**	**13317**	**12127**

notice obvious errors in the text?). The second portion (7 questions) verifies that certain subtleties of the text were understood (*What is the text about? Who is the main character of the story? Who is funding the genome sequencing? Retrieve the right sequence of steps for genome sequencing? How many base pairs are in the genome? When the novel was written, how many people had already fully sequenced their genome? Which genetic variant is associated with a high risk for Alzheimer's disease?*).

The text is considered to be overall readable (5 Very Good and 3 Good), comprehensible (8 yes, 1 not) and containing few errors (8 seldom, 1 often). The easiest comprehension questions were well handled by the readers, who all responded correctly (4 questions). However, three questions led to different answers from the readers:

- 2 readers responded incorrectly to a seemingly simple question (*Who funded the genome sequencing of Powers?*)
- The question *At the time the story was written, how many people's genomes had been sequenced?* was ambiguous since the answer could be 8 or 9 (depending on whether Powers is counted), giving rise to different responses from readers
- Only 4 of 9 readers were able to give the correct sequence of steps in the process of genome sequencing; the translated text is not unclear on this point (the errors are on the part of the readers); this mixed result may indicate a lack of interest by some readers to the most technical aspects of the text.

In short, we can say that this survey, while very limited, nevertheless demonstrates that the text (produced according to our methodology) was considered to be acceptable and rather readable by our readers (of whom 3 indicated that they read very often, 4 rather often, and 2 seldom). We also include some remarks made in the free comments:

- *"I have noticed some mistakes, some neologisms (I considered them to be neologisms and not mistranslations because they made sense)*
- *"Very fluid text and very easy reading despite precise scientific terms"*

Table 2. BLEU after tokenization and case removal on full corpus, and time measurements for each it.

MT system used	BLEU score (full corpus)	PE (block it.) time
It.1 (not adapted)	34,79	5h37mn
It.2 (tuning on Block 1)	33,13	4h45mn
It.3 (tuning on Blocks 1+2)	34,01	4h35mn

Table 3. Automatic Evaluation (BLEU) on full corpus between MT, PE and REV (revision)

Comparison	BLEU score
MT vs PE	34,01
MT vs REV	30,37
PE vs REV	79,92

- *"I found the text a little difficult because it contains complex words and it deals with an area I do not know at all."*

4.2. The views of R. Powers's french translator

To conclude this pilot study, the views of a tenth reader were solicited: the author's French translator, J-Y Pellegrin, research professor at Paris-Sorbonne. His comments are summarized here in the form of questions and answers.

Readability? *"The text you have successfully reproduces faithfully the content of the article by Powers. The readability bet is won and certain parts (in particular those which relate to the scientific aspects of the described experiment) are very convincing."*

So the MT+PE pipeline seems also efficient for obtaining quickly readable literary texts, as it is the case for other domain specific data types.

Imperfections? *"There are, of course, imperfections, clumsy expressions, and specific errors which require correction"*

Top mistakes?

- *"The most frequent defect, which affects the work of any novice translator, is the syntactic calque, where French structures the phrase differently .../... One understands, but it does not sound very French"*

- *"Another fairly common error is the loss of idiomatic French in favor of Anglicisms." Sometimes these Anglicisms can be more disturbing when flirting with Franglais*[13]*, such as translating actionable knowledge as connaissances actionnables (p. 18) instead of connaissances pratiques / utilisables."*

- *"A third defect is due to not taking into account certain cultural references .../... For example, Powers made several references to the topography of Boston that give rise to inaccuracies in the translation: "Charles River" for example (p. 12) is not une riviere but un fleuve; that is why we translate by la Charles River or simply la Charles"*

The errors mentionned above are considered as not acceptable by a professional translator of literary text. These are hard problems for computer assisted translation (move away from the syntactic calque, better handle idioms and multi-word expressions, take into account cultural references).

Could this text serve as a starting point for a professional literary translator? *"Instinctively, I am tempted to say no for now, because from his first cast the translator has reflexes that allow him to produce a cleaner text than the one you produced .../.... however, this translator would spend more than 25 hours to produce the 42 pages of 1500 characters that comprise Power's text. At a rate of 7 pages per day on average, it would take 6 eight-hour days*[14]*. If, however, I could work only from your text (while completely forgetting Powers's) and I could be guar-*

[13] Frenglish

[14] So we can consider that our computer assisted methodology accelarated the translation process by a factor 2: 25h instead of 50h

anteed that your translation contains no errors or omissions from the original, but just that it needs to be improved, made more fluid, more authentically French, things would be different and the time saved would be probably huge.".

As expected, the professional translator of literature wants to control the whole translation process. But the last part of his comment is interesting: if the meaning was guaranteed, he could concentrate on the form and limit going back and forth between source and target text. Thus, working more on quality assesment of MT and confidence estimation seems to be a promising way for future work on literary text translation.

5. CONCLUSION

5.1. Collected Data Available Online

The data in this article are available at *github.com/powersmachinetranslation/DATA*.
There one can find:

- The 545 source and target (MT, PE) segments mentioned in Table 1
- The translated and revised work (REV in Table 3), in French, that was read by a panel of 9 readers
- The results of the survey (9 readers) compiled in a spreadsheet (in French).

5.2. Comments and open questions

We presented an initial experiment of machine translation of a literary work (English text of about twenty pages). The results of an MT+PE pipeline were presented and, going beyond that, the opinions of a panel of readers and a translator were solicited. The translated text, obtained after 25 hours of human labor (a professionnal translator told us that he would have needed at least twice more) is acceptable to readers but the opinion of a professional translator is mixed. This approach suggests a methodology for rapid "low cost" translation , similar to the translation of TV series subtitles found on the web. For the author, this presents the possibility of having his work translated into more languages (several dozen instead of a handful, this short story by R. Powers has also been translated into Romanian using this same methodology).

But would the author be willing to sacrifice the quality of translation (and control over it) to enable wider dissemination of his works? For a reader who cannot read an author in the source language, this provides the ability to have faster access to an (admittedly imperfect) translation of their favorite author. For a non-native reader of the source language this provides a mechanism for assistance on the parts he or she has trouble understanding. One last thing: the title of the work *The Book of Me* has remained unchanged in the French version because no satisfactory translation was found to illustrate that the book refers both to a book but also to the DNA of the author; this paradox is a good illustration of the difficulty translating a literary work!

Thanks

Thanks to Manuela Barcan who handled the first phase of post-editing machine translations in French and Romanian during the summer of 2013. Thanks to J-Y Pellegrin, French translator of Richard Powers, for his help and open-mindedness. Thanks to R. Powers who accepted to let us work on one of his writings. Thanks to V. Bellynck, C. Boitet and L. Wang for their assistance with the SectraW post-editing tool.

References

Bbbbb AAA. 2012. The LabZZZ English to French Machine Translation System for IWSLT 2012. In *In proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT)*, dec.

AA BB. 2013. Online production of hq parallel corpora and permanent task-based evaluation of multiple mt systems. In *Workshop WPPT during MT Summit*.

Ignacio Garcia. 2011. Translating by post-editing: is it the way forward? *Journal of Machine Translation*, 25(3):217–237.

Hieu Hoang, Alexandra Birch, Chris Callison-burch, Richard Zens, Rwth Aachen, Alexandra Constantin, Marcello Federico, Nicola Bertoldi, Chris Dyer, Brooke Cowan, Wade Shen, Christine Moran, and Ondřej Bojar. 2007. Moses: Open source toolkit for statistical machine translation. In *ACL'07, Annual Meeting of the Association for Computational Linguistics*, pages 177–180, Prague, Czech Republic.

Pavel Pecina, Antonio Toral, and Josef van Genabith. 2012. Simple and effective parameter tuning for domain adaptation of statistical machine translation. In

Proceedings of the 24th International Conference on Computational Linguistics, pages 2209–2224, Mumbai, India. Coling 2012 Organizing Committee.

L Specia, N Cancedda, and M Dymetman. 2010. A dataset for assessing machine translation evaluation metrics. In *7th Conference on International Language Resources and Evaluation (LREC-2010)*, pages 3375–3378, Valetta, Malta.

Uuuuu SSSS. 2008. Sectra_w.1: an online collaborative system for evaluating, post-editing and presenting mt translation corpora. In *LREC'08, Sixth International Conference on Language Resources and Evaluation*, pages 28–30, Marrakech, Morocco.

Andreas Stolcke. 2002. Srilm: An extensible language modeling toolkit. In *ICSLP'02, 7th International Conference on Spoken Language Processing*, pages 901–904, Denver, USA.

Rob Voigt and Dan Jurafsky. 2012. Towards a literary machine translation, the role of referential cohesion. In *Computational Linguistics for Literature, Workshop at NAACL-HLT 2012*, Montreal, Canada.

Y XXX. 2012. Collection of a large database of french-english smt output corrections. In *Proceedings of the eighth international conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey, May.