Automated translation of a literary work: a pilot study

Laurent Besacier
LIG, Univ. Grenoble - Alpes
UJF - BP 53
38041 Grenoble Cedex 9, France
`laurent.besacier@imag.fr`

Lane Schwartz
Department of Linguistics
University of Illinois
Urbana, IL 61801, USA
`lanes@illinois.edu`

Abstract

Current machine translation (MT) techniques are continuously improving. In specific areas, post-editing (PE) can enable the production of high-quality translations relatively quickly. But is it feasible to translate a literary work (fiction, short story, etc) using such an MT+PE pipeline? This paper offers an initial response to this question. An essay by the American writer Richard Powers, currently not available in French, is automatically translated and post-edited and then revised by non-professional translators. 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 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.

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 (post-editors) 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 (Zhechev, 2012) show that post-editing machine translation output significantly increases productivity when compared to translating a document from scratch. This result held 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 (Zhechev, 2012) 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?

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

This work attempts to provide preliminary answers to these questions. In addition to our experimental results, we also present a new transla-

---

[1] The work of Garcia (2011) is somewhat controversial, because the manual translation without post-editing appears to have been done without allowing the translator to use any form of digital assistance, such as an electronic dicitonary.

tion (into French) of an English-language essay (*The Book of Me* by Richard Powers).

This paper is organized as follows. We begin in §2 by surveying related work in machine translation in the literary domain. In §3, we present our experimental methodology, including the choice of literary work to be translated and the machine translation, domain adaptation, and post-editing frameworks used. In §4, we present our experimental results,[2] including an assessment of translation quality using automated machine translation metrics. In §5, we attempt to assess machine translation quality beyond automated metrics, through a human assessment of the final translation; this assessment was performed by a panel of readers and by the official French translator of Richard Powers.

2   Related Work

While the idea of post-editing machine translations of scientific and technical works is nearly as old as machine translation (see, for example (Oettinger, 1954)), very little scholarship to date has examined the use of machine translation or post-editing for literary documents. The most closely related work (Voigt and Jurafsky, 2012) that we were able to identify was presented at the *ACL workshop on Computational Linguistics for Literature*[3] ; since 2012, that workshop has examined the use of NLP in the literary field. Voigt and Jurafsky (2012) 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 feasible or not.

For the purposes of this paper, we now define what constitutes a literary text. 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 such 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).

3   Methodology

For this study, we follow a variant of the post-editing methodology established by Potet et al. (2012). In that work, 12,000 post-edited segments (equivalent to a book of about 500 pages) in the news domain were collected through crowdsourcing, resulting in one of the largest freely available corpora of post-edited machine translations.[4] It is, for example, three times larger than that collected by Specia et al. (2010), a well known benchmark in the field.

Following Potet et al. (2012), we divide the document to be translated into three equal parts. A translation/post-edition/adaptation loop was applied to the three blocks of text according to the following process:

- The first third of the document was translated from English to French using Moses (Hoang et al., 2007), a state-of-the-art phrase-based machine translation system. This machine translation output was then post-edited.

- The post-edited data from the third of the document was used to train an updated domain-adapted English-French MT system. Given the small amount of post-edited data, adaptation at this point consisted only 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.

Our methodology differs in two important ways from Potet et al. (2012). First, our study makes use of only one post-editor, and does not use crowd-sourcing to collect data. Second, once the post-editing was completed, the final text was revised: first by the post-editor and then by another reviewer. The reviewer was a native French speaker with a good knowledge of the English language. The times taken to post-edit and revise were recorded.

3.1 Choice of literary document

To test the feasibility of using machine translation and post-editing to translate a literary work, we began by selecting an essay written in English which had not yet been translated into French. The choice of text was guided by the following factors:

(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) In his writing, Powers often explores the effects of modern science and technology, and in some ways his writings contain commonalities with scientific and technical texts. We hypothesized that this characteristic may somewhat reduce the gap between translation of scientific and literary texts.

Via his French translator (Jean-Yves Pellegrin), Powers was informed by e-mail of our approach, and he gave his consent as well as his feeling on this project (Richard 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 selected an essay by Powers, entitled *The Book of Me*, originally published in GQ magazine.[6] The essay is a first-person narrative set in 2008, in which Powers describes the process by which he became the ninth person in the world to see his genome fully sequenced. 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 level, never puts aside his poetic ambition, his humour and his fascination for the impact of science and technology on the society.

3.2 MT system used

Our machine translation system is a phrase-based system using the Moses toolkit (Hoang et al., 2007). Our system is trained using the data provided in the IWSLT machine translation evaluation campaign (Federico et al., 2012), representing a cumulative total of about 25M sentences:

- *news-c*: version 7 of the News-Commentary corpus,

- *europarl*: version 7 of the Europarl corpus[7] (Koehn, 2005),

- *un*: the United-nations corpus,[8]

- *eu-const*: corpus which is freely available (Tiedemann, 2009),

- *dgt-tm*: DGT Multilingual Translation Memory of the Acquis Communautaire (Steinberger et al., 2012),

- *pct*: corpus of Parallel Patent Applications[9],

- *gigaword*: 5M sentences extracted from the Gigaword corpus; after cleaning, the whole Gigaword corpus was sorted at sentence level according to the sum of perplexities of the source (English) and the target (French) based on two French and English pretrained language models. Finally, the 5M subset was obtained after filtering out the whole Gigaword corpus with a cut-off limit of 300 (ppl). This leads to a subset of 5M aligned sentences.

Prior to training the translation and language models, various pre-processing steps are performed on the training data. We begin by filtering out

---

[5]http://en.wikipedia.org/wiki/Richard_
Powers
[6]http://www.gq.com/
news-politics/big-issues/200810/
richard-powers-genome-sequence

[7]http://www.statmt.org/europarl/
[8]http://www.euromatrixplus.net/multi-un/
[9]http://www.wipo.int/patentscope/en/
data/pdf/wipo-coppatechnicalDocumentation.
pdf

badly aligned sentences (using several heuristics), filtering out empty sentences, and sentences having more than 50 words. Punctuation is normalized, and we tokenize the training data, applying specific grammar-based rules for the French tokenization. Spelling correction is applied to both source and target side, and certain words (such as *coeur*) are normalized. Abbreviations and clitics are disambiguated. Various additional cleaning steps (as described in the list above) were applied to the Gigaword corpus. Many heuristics (rules) were used in order to keep only good quality bi-texts.

From this data, we train three distinct translation models on various subsets of the parallel data (*ted; news-c+europarl+un+eu-const+dgt-tm+pct; gigaword5M*). The French part of the same corpus is used for language model training, with the addition of the *news-shuffle* corpus provided as part of the WMT 2012 campaign (Callison-Burch et al., 2012). 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 weights for the final machine translation system are optimized using the data from the English-French MT task of IWSLT 2012. The system obtains BLEU (Papineni et al., 2002) scores of 36.88 and 37.58 on the IWSLT tst2011 and test2012 corpora, respectively (BLEU evaluated with case and punctuation).

The training data used is out-of-domain for the task of literary translation, and as such is clearly not ideal for translating literary texts. In future work, it would be desirable to at least collect literary texts in French to adapt the target language model, and if possible gain access to other works and translations of the same author. Additionally, in future work we may examine the use of real-time translation model adaptation, such as Denkowski et al. (2014).

3.3 Post-editing

We use the SECTra_w.1 post-editing interface of Huynh et al. (2008). This tool also forms the foundation that gave rise to the interactive Multilingual Access Gateway (iMAG) framework for enabling multilingual website access, with incremental improvement and quality control of the translations. It has been used for many projects (Wang and Boitet, 2013), including translation of the EOLLS encyclopedia, as well as multilingual access to dozens of websites (80 demonstrations, 4 industrial contracts, 10 target languages, 820k post-edited segments).

Figure 1 shows the post-editing interface in advanced mode. In advanced mode, multiple automatic translations of each source segment (for example, from Google, Moses, etc.) can be displayed and corrected. For this experiment, the output of our Moses system was prioritized when displaying segment translations.

Post-editing was done by a (non-English native) student in translation studies at Université Grenoble Alpes.

4 Experimental results

4.1 Corpus and post-editing statistics

The test data, *The Book of Me* (see §3.1), is made up of 545 segments comprising 10,731 words. This data was divided into three equal blocks. We apply machine translation and post-editing to the data, as described in §3.2 and §3.3.

Table 1 summarizes the number of source and target (MT or PE) 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. The post-editing results reported in Table 1 are obtained after each iteration of the process; the last stage of revision is thus not taken into account at this stage.

4.2 Performance of the MT system

Table 2 summarizes machine translation performance, as measured by BLEU (Papineni et al., 2002), 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 in-

Figure 1: Post-editing interface in advanced mode

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

Table 1: Number of words in each block of the English source corpus, French machine translation (MT), and French post-edited machine translation (PE).

creasing productivity as the post-editor adapts to the task). In the end, the total PE time is estimated at about 15 hours.

4.3 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 original post-editor (4 hours) and one by a second French-English bilingual serving as a reviewer (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 (PE - 15 hours of work) and the revised version (REV - 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. This can be observed by computing BLEU between the post-edited text before and after revising; the result is a BLEU score

of 79.92, indicating very high similarity between the two versions. So, post-editing and revising are very different tasks. This is illustrated by the numbers of Table 3: MT and PE are highly dissimilar (post-editor corrects a lot of MT errors) while PE and REV are similar (revision probably focuses more on important details for readability and style). 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 — see footnote 2 on page 2).

5 Human evaluation of post-edited MT

5.1 The views of readers on the post-edited translation

Nine French readers agreed to read the final translated work and answer an online questionnaire. The full survey is available on *fluidsurveys.com*.[10]. A pdf version of the test results and a spreadsheet file containing the results of the survey are also made avail-

---

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

118

MT system used	BLEU score (full corpus)	PE (block it.) time
Iteration 1 (not adapted)	34.79	5h 37mn
Iteration 2 (tuning on Block 1)	33.13	4h 45mn
Iteration 3 (tuning on Blocks 1+2)	34.01	4h 35mn

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

Comparison	BLEU score
MT vs PE	34.01
MT vs REV	30.37
PE vs REV	79.92

Table 3: Automatic Evaluation (BLEU) on full corpus between unedited machine translation (MT), post-edited machine translation (PE), and revised post-edited machine translation (REV).

able on *github* (see footnote 2 on page 2).

After three questions to better understand the profile of the participant (*How old are you? Do you frequently read? If yes, what is your favorite genre?*), the first portion asks readers five questions 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 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?*

- *Chronologically sort the sequence of steps involved in 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 in 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"*

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

5.2 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, Jean-Yves Pellegrin, research professor at Paris-Sorbonne University. 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,[11] such as translating 'actionable knowledge' as 'connaissances actionnables' (p. 18) instead of 'connaissances pratiques / utilisables'."*

---

[11]Frenglish

- *"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 mentioned 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. If, however, I could work only from your text (while completely forgetting Powers's) and I could be guaranteed 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 were 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. Based on the translation speed rates provided by Pellegrin, we can estimate the time savings of our technique. Our computer-assisted methodology can be said to have accelerated the translation factor by a factor of 2 — our process took roughly 25 hours, compared to the 50 hours estimated for a professional literary translation.

6 Conclusion

6.1 Collected Data Available Online

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

- The 545 English source and French 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)

6.2 Comments and open questions

We presented an initial experiment of machine translation of a literary work (an 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 professional translator told us that he would have needed at least twice that much time) 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 of the literary text, this presents the possibility of having his work translated into more languages (several dozen instead of a handful, this short story by Richard 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 author refers both to a book but also to his DNA; this paradox is a good illustration of the difficulty translating a literary work!

References

Chris Callison-Burch, Philipp Koehn, Christof Monz, Matt Post, Radu Soricut, and Lucia Specia. 2012. Findings of the 2012 Workshop on Statistical Machine Translation. In *Proceedings of the Seventh Workshop on Statistical Machine Translation*, pages 10–51, Montréal, Canada, June. Association for Computational Linguistics.

Michael Denkowski, Alon Lavie, Isabel Lacruz, and Chris Dyer. 2014. Real time adaptive machine translation for post-editing with cdec and TransCenter. In *Proceedings of the EACL 2014 Workshop on Humans and Computer-assisted Translation*, pages 72–77, Gothenburg, Sweden, April. Association for Computational Linguistics.

Marcello Federico, Mauro Cettolo, Luisa Bentivogli, Michael Paul, and Sebastian Stüker. 2012. Overview of the IWSLT 2012 evaluation campaign. In *In proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT)*, December.

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.

Cong-Phap Huynh, Christian Boitet, and Hervé Blanchon. 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.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *Proceedings of the Tenth Machine Translation Summit*, Phuket, Thailand, September.

Anthony Oettinger. 1954. *A Study for the Design of an Automatic Dictionary*. Ph.D. thesis, Harvard University.

Kishore Papineni, Salim Roukos, Todd Ward, and Zhu Wei-Jing. 2002. BLEU: A method for automatic evaluation of machine translation. In *ACL'02, 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, PA, USA.

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.

Marion Potet, Emmanuelle Esperança-Rodier, Laurent Besacier, and Hervé Blanchon. 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.

Lucia Specia, Nicola Cancedda, and Marc 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.

Ralf Steinberger, Andreas Eisele, Szymon Klocek, Spyridon Pilos, and Patrick Schlüter. 2012. DGT-TM: A freely available translation memory in 22 languages. In *LREC 2012*, Istambul, Turkey.

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

Jörg Tiedemann. 2009. News from OPUS - a collection of multilingual parallel corpora with tools and interfaces. In *Proceedings of Recent Advances in Natural Language Processing*.

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.

Lingxiao Wang and Christian Boitet. 2013. Online production of HQ parallel corpora and permanent task-based evaluation of multiple MT systems. In *Proceedings of the MT Summit XIV Workshop on Post-editing Technology and Practice*.

Ventsislav Zhechev. 2012. Machine translation infrastructure and post-editing performance at Autodesk. In *AMTA'12, Conference of the Association for Machine Translation in the Americas*, San Diego, USA.