

Monolingual Post-Editing by a Domain Expert is Highly Effective for Translation Triage

Lane Schwartz

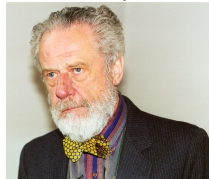
University of Illinois at Urbana-Champaign
lanes@illinois.edu

26 October 2014

The Proper Place of Men and Machines in Language Translation

“ *The world is badly in need of translators. Almost nobody denies this. The number of pairs of languages between which translations must be made and the number and types of documents involved is constantly increasing.* ”

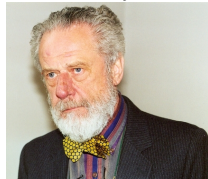
Martin Kay (1980)



The Proper Place of Men and Machines in Language Translation

“ *But we are fortunate to be children of the age of computers and it is to them that we naturally turn. A computer is a device that can be used to magnify human productivity.* ”

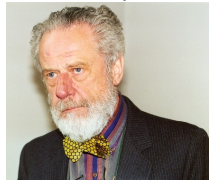
Martin Kay (1980)



The Proper Place of Men and Machines in Language Translation

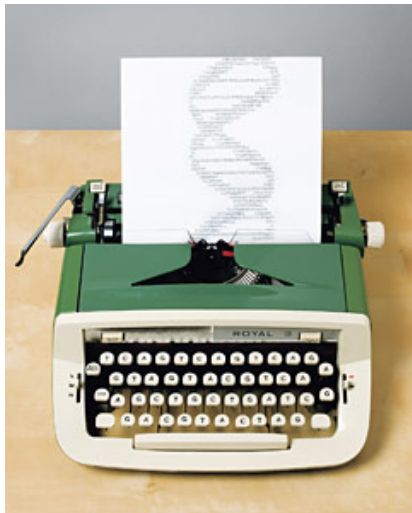
“ *Properly used, it does not dehumanize by imposing its own Orwellian stamp on the products of the human spirit and the dignity of human labor but, by taking over what is mechanical and routine, it frees human beings for what is essentially human.* ”

Martin Kay (1980)



A Post-Editing Case Study

The Book of Me, by Richard Powers



Post-Editing

●●○○○○○○○

Monolingual Post-Editing by a Domain Expert

○○○○○○○○○○

Evaluation

○○○○○○○○○○○○

References

The Book of Me, by Richard Powers



Post-Editing

○○●○○○○○

Monolingual Post-Editing by a Domain Expert

○○○○○○○○○○

Evaluation

○○○○○○○○○○○○

References

The Book of Me, by Richard Powers



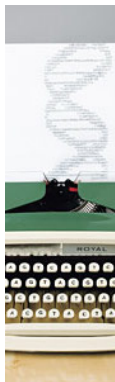
The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



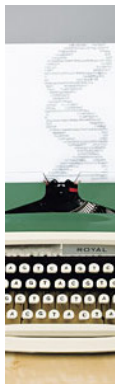
The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



The Book of Me, by Richard Powers



- 10,731 English source words
- 12,127 post-edited French target words
- 15 hours of post-editing
- 10 hours of revision

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

Readability? *The readability bet is won and certain passages (in particular those which relate to the scientific aspects of the described experiment) are very convincing.*

Could this serve as a starting point for a prof. literary translation?

"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 .../...."

Could this serve as a starting point for a prof. literary translation?

At a rate of 7 pages per day on average, it would take 6 eight-hour days

Could this serve as a starting point for a prof. literary translation?

If, however, I could work only from your text (while completely forgetting Powers) 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 considerable.”

Monolingual Post-Editing

Bilingual Translator



Post-Editing



Monolingual Post-Editing by a Domain Expert



Evaluation



References



Bilingual Translator++



Post-Editing
○○○○○○○○

Monolingual Post-Editing by a Domain Expert
○●○○○○○○

Evaluation
○○○○○○○○○○

References

Monolingual Post-Editing

(Callison-Burch, 2005)

Linear B: Post-Editing Translations

<http://www.linearb.co.uk/nist2005/XIA20050105.0130.html>

There are 10 sentences in this article. Please edit their translations so that they make sense and are good English.

Please edit this translation:

Emphasize that Turkey would reject any pressures to press for recognition of the Cyprus

Please edit this translation:

Ankara Jan 12 (AFP) - The head of the Turkish Government Tayyib Rajab Wednesday that Ankara would reject any pressures had exercised by the European Union to press for recognition of Cyprus, in two weeks ago Summit of Heads of State and Government of the European Union who will determine the opening of accession negotiations with Ankara or not.

Please edit this translation:

The station in the Te television that "the EU can do that by imposing new conditions on us Cyprus.

Monolingual Post-Editing

(Mitchell et al., 2013)

The screenshot shows a software window titled "Bearbeiten" with a close button in the top right corner. The window is divided into several sections:

- Klicken Sie auf den Text um diesen zu bearbeiten:** This section contains a text area with a highlighted segment: "Lösung für Windows XP nicht mehr Windows Updates von geschieft - ComboFix.exe". Below the highlight is a paragraph of German text: "- Website www.bleepingcomputer.com/download/combofix / und laden Sie ComboFix.exe ausführen. Sehr viel Geduld. Es geht aber wie 100 Phasen und Neustarts, aber es funktioniert. Ich habe für mich! ----- Benutzer NICHT combofix ausgebildet ausführen, ohne Hilfestellung, es ist ein leistungsstarkes Tool, das kann zu Problemen mit dem System Selbst die Schöpfer Staat und nicht, um sie auszuführen ich hatte, System wieder nach der Ausführung combofix und andere Tools, aber ich bin in den richtigen Bereich zu tun. Und es ist nicht zu 100 Phasen, Thread Ersteller kann nicht einmal das richtig."
- Aktueller Satz zum Bearbeiten:** This section displays the current sentence being edited: "Lösung für Windows XP nicht mehr Windows Updates von geschieft - ComboFix.exe". Below the text are navigation icons: a left arrow, a right arrow, and a keyboard icon labeled "ABC".
- Zurück Weiter:** Two buttons for navigating between sentences.
- Kommentare ?** A large empty text area for adding comments.
- Bottom Bar:** Contains a help icon (question mark in a circle) on the left and three action buttons: "Aufgabe fertigstellen", "Richtlinien", and "für später speichern".

Monolingual Translation Options

(Callison-Burch, 2005)

Linear B: Visualization Tool

<http://www.linearb.co.uk/nist2005/segment1.html>

اردوغان	يؤكد	بأن	تركيا	سترفض	اي	ضغوطات	لحثها	على	الاعتراف	بقبرص
		that turkey that turkey is turkey that turkey had that turkey will						to recognize recognition recognition of to recognize the the recognition		
	confirm that the confirm that affirm that confirms that who assert that our						urge it to to urge them to and urge it to order to urge it to urging			
ardogan ardogan ,	affirms stresses emphasize confirms emphasizes	that that the to , that that ,	turkey of turkey by turkey , turkey turkey ' s	reject will reject refuse not accept would be	any no of any i . e . meaning	pressure to pressure up to pressure	urge it urge and urge it urging urge them to	on to on the for to the	recognition recognition of recognized recognition of the	cyprus cyprus , to cyprus on cyprus cyprus and

Monolingual Translation Options

(Koehn, 2010)

وكان	مجلس	التواب	الاميركي	اعتمد	الخميس	قانونا	يطالب	بسحب	القوات	المقاتلة	الاميركية	من	العراق	في	موعد	اقصاه	الاول	من	تيسان @/ @	ايوزل
the	the us house of representatives	adopted	thursday	legally	calls for the withdrawal of	combat troops	us	iraq	in	no later than	the first	from	april							
the us house of representatives	the	thursday ,	law	the fighting forces	the us	from iraq	in	the latest	the first of	april										
the us house	adopted the	thu	the legally	fighting forces	us	from iraq	in	i	april											
it was	us house of representatives	was adopted	thursday , the	the law	demands withdrawal of troops	fighter	the us	no later than	first	on april										
he was	the us house	adopted by	thursday 's	a law	calls for withdrawal of	combat forces	of	in the	not later than	first of										
he	us house	adopted by the	on thursday	a legally	calls for the withdrawal	forces	the fighter	from												
earlier ,	us	adopted a	on thursday ,	by law	demands the withdrawal of	troops	iraq													
was		, was adopted	thursday the	legally ,	demands withdrawal of		of the													
it was the		adopted ,	thu ,	the legal	calls for withdrawal		from iraq	in the												
earlier , the		adopted , the	thursday , a	legally @-@	demands the withdrawal		the american													
2008 ,	متحدياً	مرة	جديدة	الرئيس	جورج	بوش	الذي	يعارض	أي تحديد	موعد										
2008 ,	defying	once	new	president george w. bush	who opposes	no date has been set for the														
the 2008	defiant	once again		president george bush	, who opposes	no date has been set for														
2008	challenging	again	the new		, who opposes	no date has been set														
	a defiant	the first			, who opposes the	a date .														
	in defiance of	once again ,			, who opposes	date .														
	, challenging	once again the		president george bush , who	opposed to setting any	the date of the														
	, in defiance	for the first time	a new	president george w. bush 's	which opposes	no date														
in 2008	defying the	again		us president george w. bush	opposed to	any														
	challenging the	time			who opposes the	date of														
	, defying	once again , the			opposes	date														

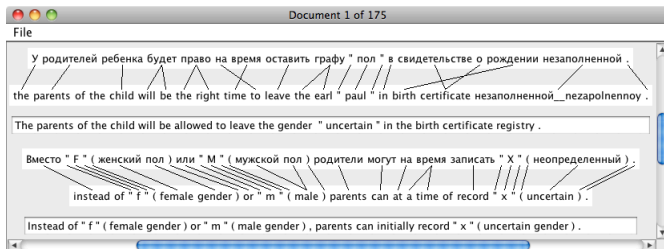
Monolingual Post-Editing with Alignments

(Albrecht et al., 2009)

The screenshot displays the 'The Chinese Room' software interface. The main window is titled 'The Chinese Room' and has a menu bar with 'Edit' and 'Style'. Below the menu bar are navigation buttons: 'Sent: Prev Next Doc: Prev Next Pop'. On the right side, there are checkboxes for 'Show Parse' and 'Show Alignments', both of which are checked. A 'Select Translation' dialog box is open on the right, showing search results for the phrase '今年前'. The results list several entries with Chinese characters and English descriptions, such as '今年前 3' (first three months of this year) and '今年前 6' (first 6 months of this year). The main workspace is divided into several sections. On the left, there are input boxes for 'Two years ago', 'this month', 'Guangdong', 'new high', and 'tech'. In the center, there are vertical alignment lines with Chinese characters and brackets indicating relationships between the input and the target text. On the right, there are target text boxes with corresponding Chinese characters and English descriptions, such as 'this year', 'today, modern, present, current, this, now', 'year', 'before, in front, a former, previous, of front', 'both, two, ounce, a few, tael', 'moon, month', 'wide, numerous, to east', 'high, tall', 'new, newly', 'skill', and 'to reproduce, to pr...'. The bottom status bar shows '100%'.

Monolingual Post-Editing with Alignments

(Schwartz et al., 2014)

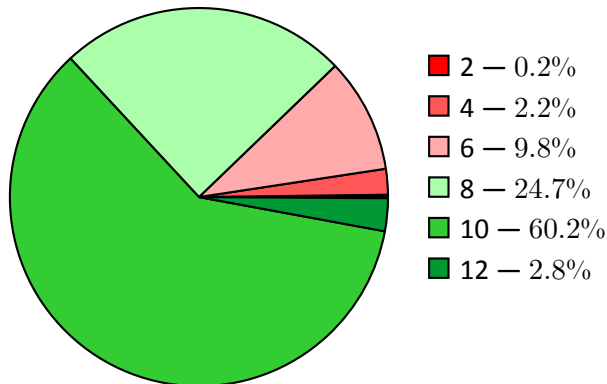


WMT14 Translation Adequacy Guidelines

12	The postedited translation is superior to the reference translation
10	The meaning of the Russian sentence is fully conveyed in the English translation
8	Most of the meaning of the Russian sentence is conveyed in the English translation
6	The English translation misunderstands the Russian sentence in a major way, or has many small mistakes
4	Very little information from the Russian sentence is conveyed in the English translation
2	The English translation makes no sense at all

Adequacy evaluation guidelines for bilingual human judges (Schwartz et al., 2014), adapted from Albrecht et al. (2009).

WMT14 Translation Adequacy Results



Percentage of evaluated sentences judged to be in each category by a bilingual judge (Schwartz et al., 2014).

Evaluation

Experimental Setup: Machine Translation Systems



Post-Editing

○○○○○○○○

Monolingual Post-Editing by a Domain Expert

○○○○○○○○○

Evaluation

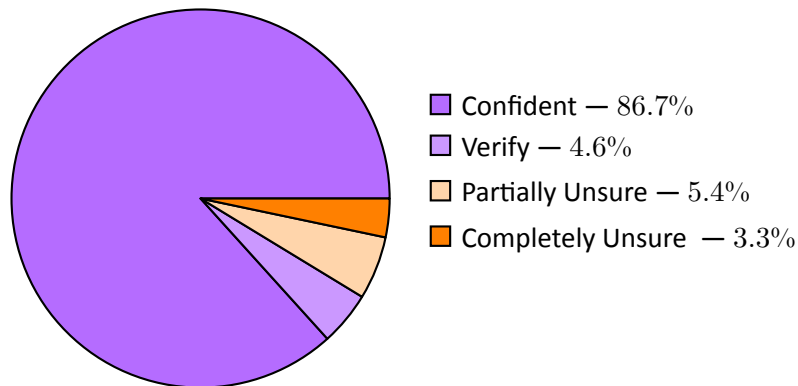
●○○○○○○○○○○

References

Post-Editor Confidence Guidelines

Confident	The monolingual post-editor is confident that the post-edited translation conveys the meaning of the French sentence
Verify	The monolingual post-editor believes that the post-edited translation conveys the meaning of the French sentence, but the translation should be verified by a bilingual post-editor
Partially unsure	The monolingual post-editor is not confident that a specific portion of the post-edited translation is correct; that section should be handled by a bilingual post-editor
Completely unsure	The entire sentence should be handled by a bilingual post-editor

Post-Editor Confidence Results



Selected Segment: Unsure

Source: Untel Trucmuche , Unetelle Machinchose

Google Trucmuche Doe, Jane Doe Machinchose

Systran So-and-so Thingy, Machinchose So-and-so

Moses so-and-so trucmuche unetelle machinchose ,

Monolingual Post-Editor: Untel Trucmuche , Unetelle Machinchose

Bilingual Post-Editor: First Author, Second Author

Source: Untel Trucmuche , Unetelle Machinchose

Google Trucmuche Doe, Jane Doe Machinchose

Systran So-and-so Thingy, Machinchose So-and-so

Moses so-and-so trucmuche unetelle machinchose ,

Monolingual Post-Editor: Untel Trucmuche , Unetelle Machinchose

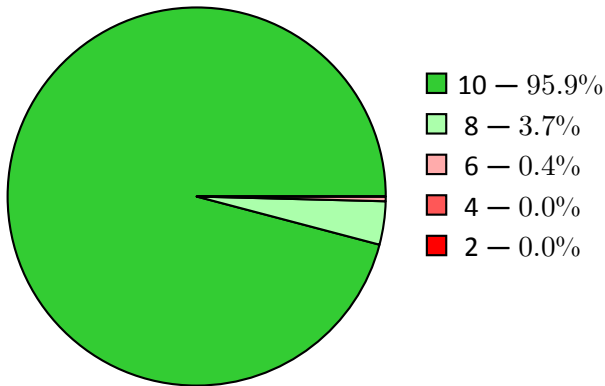
Bilingual Post-Editor: First Author, Second Author

Translation Adequacy Guidelines

10	The meaning of the French sentence is fully conveyed in the English translation
8	Most of the meaning of the French sentence is conveyed in the English translation
6	The English translation misunderstands the French sentence in a major way, or has many small mistakes
4	Very little information from the French sentence is conveyed in the English translation
2	The English translation makes no sense at all

Adequacy evaluation guidelines for bilingual human judges, adapted from Albrecht et al. (2009).

Translation Adequacy Results



Percentage of 241 evaluated segments judged to be in each translation adequacy category by a bilingual judge.

Post-Editor Confidence vs Translation Adequacy

		Evaluation Category				
		2	4	6	8	10
Post-Editor Confidence	Completely unsure	0	0	0	0	8
	Partially Unsure	0	0	0	2	11
	Verify	0	0	0	1	10
	Confident	0	0	1	6	202

For each of the 241 evaluated sentences, the adequacy category assigned by a bilingual judge, along with confidence assigned by the post-editor.

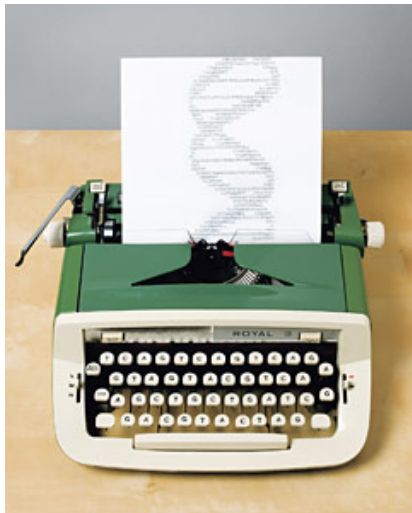
Post-Editor Confidence vs Translation Adequacy

		Evaluation Category				
		2	4	6	8	10
Post-Editor Confidence	Completely unsure	0	0	0	0	8
	Partially Unsure	0	0	0	2	11
	Verify	0	0	0	1	10
	Confident	0	0	1	6	202

For each of the 241 evaluated sentences, the adequacy category assigned by a bilingual judge, along with confidence assigned by the post-editor.

That Case Study, Revisited

The Book of Me, by Richard Powers



Post-Editing

○○○○○○○○

Monolingual Post-Editing by a Domain Expert

○○○○○○○○○○

Evaluation

○○○○○○○○●○○

References

Traduction automatisée d'une oeuvre littéraire: une étude pilote

Laurent Besacier¹

(1) LIG, Université de Grenoble, UJF - BP 53, 38041 Grenoble Cedex 9
laurent.besacier@imag.fr

Résumé. Les techniques actuelles de traduction automatique (TA) permettent de produire des traductions dont la qualité ne cesse de croître. Dans des domaines spécifiques, la post-édition (PE) de traductions automatiques permet, par ailleurs, d'obtenir des traductions de qualité relativement rapidement. Mais un tel pipeline (TA+PE) est-il envisageable pour traduire une oeuvre littéraire ? Cet article propose une ébauche de réponse à cette question. Un essai de l'auteur américain Richard Powers, encore non disponible en français, est traduit automatiquement puis post-édité et révisé par des traducteurs non-professionnels. La plateforme de post-édition du LIG utilisée permet de lire et éditer l'oeuvre traduite en français continuellement, suggérant (pour le futur) une communauté de lecteurs-réviseurs qui améliorent en continu les traductions de leur auteur favori. En plus de la présentation des résultats d'évaluation expérimentale du pipeline TA+PE (système de TA utilisé, scores automatiques), nous discutons également la qualité de la traduction produite du point de vue d'un panel de lecteurs (ayant lu la traduction en français, puis répondu à une enquête). Enfin, quelques remarques du traducteur français de R. Powers, sollicité à cette occasion, sont présentées à la fin de cet article.

- In 21^{ème} *Traitement Automatique des Langages Naturelles*
- <http://aclweb.org/anthology/F/F14/F14-2001.pdf>

Bilingual Translator++



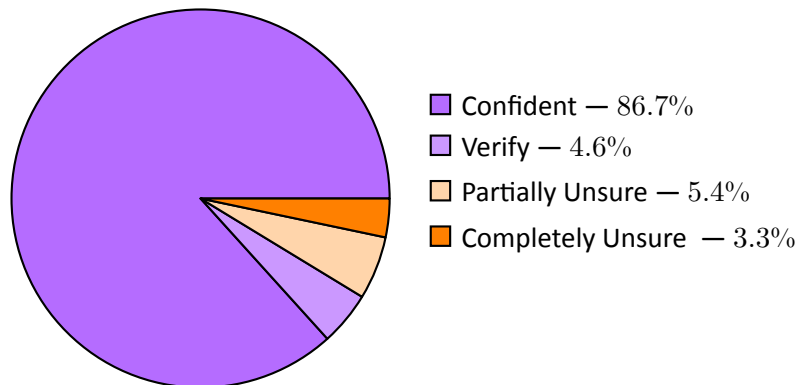
Post-Editing
○○○○○○○○

Monolingual Post-Editing by a Domain Expert
○○○○○○○○○○

Evaluation
○○○○○○○○○○●○

References

Post-Editor Confidence Results



Traduction automatisée d'une oeuvre littéraire : une étude pilote

Laurent Besacier¹

(1) LIG, Université de Grenoble, UJF - BP 53, 38041 Grenoble Cedex 9
laurent.besacier@imag.fr

Résumé. Les techniques actuelles de traduction automatique (TA) permettent de produire des traductions dont la qualité ne cesse de croître. Dans des domaines spécifiques, la post-édition (PE) de traductions automatiques permet, par ailleurs, d'obtenir des traductions de qualité relativement rapidement. Mais un tel pipeline (TA + PE) est-il envisageable pour traduire une oeuvre littéraire ? Cet article propose une ébauche de réponse à cette question. Un essai de l'auteur américain Richard Powers, encore non disponible en français, est traduit automatiquement puis post-édité et révisé par des traducteurs non-professionnels. La plateforme de post-édition du LIG utilisée permet de lire et éditer l'oeuvre traduite en français continuellement, suggérant (pour le futur) une communauté de lecteurs-réviseurs qui améliorent en continu les traductions de leur auteur favori. En plus de la présentation des résultats d'évaluation expérimentale du pipeline TA + PE (système de TA utilisé, scores automatiques), nous discutons également la qualité de la traduction produite du point de vue d'un panel de lecteurs (ayant lu la traduction en français, puis répondu à une enquête). Enfin, quelques remarques du traducteur français de R. Powers, sollicité à cette occasion, sont présentées à la fin de cet article.

Abstract. Current machine translation (MT) techniques continue to improve. In specific fields, post-editing (PE) of automatic translations can produce high-quality translations relatively quickly. But can such a (MT + EP) pipeline be used to translate a literary work ? This paper attempts a preliminary answer to this question. A short story by the American author Richard Powers, which is currently unavailable in French, is automatically translated and then post-edited and revised by non-professional translators. The LIG post-editing platform is used to view and edit the short story, potentially enabling a future community of readers and editors to continuously improve translations of their favorite author. In addition to presenting experimental evaluation results (MT system details and automatic evaluation results) for the MT+PE pipeline, we also discuss the quality of the post-edited translation output from the perspective of a panel of readers (who read the translation in French, and then responded to a survey). Finally, some remarks of the French translator of R. Powers, solicited on this occasion, are presented at the end of this article.

Mots-clés : Mots-clés : traduction automatique, TA, oeuvre littéraire, post-édition. .

Keywords : Keywords : machine translation, MT, literature, post-editing. .

1 Introduction

La tâche de post-édition consiste à éditer la sortie textuelle produite, le plus souvent par une machine (système de traduction automatique, de reconnaissance optique de caractères, de transcription automatique, etc.) en vue de l'améliorer. Dans le domaine de la traduction de documents, où les systèmes automatiques sont utilisés, le processus suivant est généralement utilisé : le système de TA produit des traductions brutes qui sont manuellement post-éditées par des traducteurs professionnels formés (post-éditeurs) corrigeant les erreurs de traduction.

De nombreuses études ont montré les avantages de l'utilisation combinée traduction automatique + post-édition manuelle (TA + PE) dans un flux de diffusion. Par exemple, Garcia (2011) a montré que même si la post-édition de sorties brutes de TA ne conduit pas toujours à une amélioration en termes de productivité, elle permet de produire de meilleures traductions par rapport à des traductions manuelles directement issues du texte source¹. Autodesk a aussi réalisé une expérience pour tester si l'utilisation de la TA permettrait d'améliorer la productivité des traducteurs. Dans cette expérience, les résultats² montrent que la post-édition des sorties issues d'un système automatique augmente significativement la productivité par rapport à une traduction partant "de zéro". Ceci est vérifié quelle que soit la paire de langues, l'expérience du traducteur et sa préférence exprimée concernant la méthode (post-éditer des sorties automatiques vs traduire à partir de zéro).

1. le travail de Garcia (2011) est cependant sujet à controverses car la traduction manuelle sans post-édition semble avoir été faite sans autoriser l'utilisation d'aide numérique pour assister le traducteur (même pas un dictionnaire électronique)

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

Ces résultats mesurés en milieu académique (Garcia (2011)) ou industriel (Autodesk) sur des processus de traduction dans des domaines spécialisés, nous conduisent à poser la question suivante : quelle serait la valeur d'un tel processus (TA + PE) appliqué sur la traduction d'une oeuvre littéraire ? Combien de temps faut-il pour traduire une oeuvre de dix mille mots ? La traduction obtenue est-elle acceptable pour des lecteurs ? **Qu'en penserait le traducteur attiré de l'auteur considéré ?** Une traduction "low cost" produite par des communautés de fans (comme c'est le cas pour les séries TV) est-elle envisageable pour des romans ou des nouvelles ? Cet article tente d'apporter de premières réponses (certes très partielles) à ces questions. **En plus de cette étude préliminaire, il restera au moins de ce travail la traduction inédite d'une nouvelle, ou plutôt d'un essai (*The Book of Me* de Richard Powers).**

Le suite de l'article présente notre méthodologie générale, le choix de l'oeuvre et les systèmes de TA et de PE (tous deux issus du LIG) utilisés. Les données expérimentales collectées (et rendues disponibles à la communauté sur *github*) sont ensuite décrites dans la section 3. **Dans la partie 4, nous tentons d'aller au delà d'une tâche de TA adaptée au domaine en sollicitant le point de vue d'un (petit) panel de lecteurs et l'avis du traducteur attiré de R. Powers.** Pour finir, la partie 5 conclut ce travail préliminaire.

2 Traduction automatisée d'une oeuvre littéraire

2.1 Remarques préliminaires

Le format court de cet article ne permet pas de proposer une étude bibliographique exhaustive sur la traduction assistée de documents littéraires ou, plus généralement, sur l'utilisation du TALN dans le domaine littéraire. Cependant, on peut noter que c'est un champ de recherche assez actif : un atelier y est consacré depuis 2012 : *ACL workshop on Computational Linguistics for Literature*³. **On peut aussi mentionner la tenue récente, en France, de la *Journée d'étude sur la traduction de textes littéraires* organisée par l'ENS en 2013.** Cependant, à notre connaissance, cette étude est la première sur la traduction automatisée d'une oeuvre littéraire. Il conviendrait également de définir ce qu'on appelle un texte littéraire. En ce qui nous concerne, nous incluons dans cette catégorie (notre définition est sans doute trop restrictive) toutes les oeuvres de fiction ou autobiographiques écrites sous la forme de romans ou de nouvelles. **Dans ces textes, l'auteur exprime sa vision du monde, de son époque et de la vie en général tout en utilisant des *procédés littéraires* et une *technique d'écriture* (forme) lui permettant de créer des effets à l'aide du langage ou de faire passer des messages (explicites ou sous-entendus).**

2.2 Choix de l'oeuvre

Le choix de l'oeuvre a été guidé par le fait que (a) nous avons un contact avec le traducteur français de l'auteur américain Richard Powers⁴ (auteur notamment du roman *La chambre aux échos* qui a remporté le National Book Award et a été finaliste du Prix Pulitzer) (b) R. Powers publie des romans souvent en lien avec la science et, d'une certaine manière, ses textes contiennent quelques points communs avec des textes scientifiques ou techniques (ce qui réduit peut-être un peu le fossé entre traduction de textes scientifiques et littéraires ?).

Par l'intermédiaire de son traducteur français (J-Y Pellegrin), R. Powers a été informé, par e-mail, de notre démarche et il a donné son accord. Nous avons ensuite choisi une oeuvre, non encore traduite en français, intitulée *The Book of Me* et publiée pour la première fois dans les colonnes du magazine GQ⁵. C'est un texte narratif, raconté à la première personne, où l'auteur est le personnage principal de l'histoire. Il y raconte l'année 2008 au cours de laquelle il est devenu la neuvième personne au monde à voir son génome entièrement séquencé. **Bien que le thème soit la génétique et malgré le style simple et clinique employé par l'auteur, *The Book of Me* est bien une oeuvre littéraire où l'auteur, qui enseigne la technique narrative à l'université, ne met jamais de côté son ambition poétique, son humour et sa fascination pour l'irrationnel.**

2.3 Méthodologie générale

La traduction de textes par cascade TA + PE (traduction automatique puis post-édition) a déjà été évaluée au LIG sur des données journalistiques. En 2012, la collecte de post-éditions manuelles de 12 000 énoncés (équivalente à un livre de 500 pages environ) a été réalisée par Potet et al. (2012). Ce corpus collecté via *crowdsourcing* et disponible en ligne⁶, est un des plus gros corpus existant concernant la post-édition de données issues d'un système de traduction automatique appliqué sur des données libres de droit. **Il est par exemple trois fois plus important que celui collecté par Specia et al. (2010) qui fait référence dans le domaine.** La même méthodologie a été appliquée sur l'oeuvre littéraire choisie mais (différence importante) il n'y avait qu'un seul post-éditeur (pas de *crowdsourcing*) et le texte postédité a été ensuite révisé.

3. sites.google.com/site/clfl2014a/

4. fr.wikipedia.org/wiki/Richard_Powers

5. www.gq.com/news-politics/big-issues/200810/richard-powers-genome-sequence

6. www-clips.imag.fr/geod/User/marion.potet/index.php?page=download

Plus précisément, le corpus a été divisé en trois parties égales. Une boucle *traduction-postédition-adaptation* est appliquée sur les trois blocs de texte selon le processus suivant :

- Après traduction automatique du premier tiers (par un système de TA probabiliste initial non adapté à la tâche), les sorties traduites sont révisées par une étudiante en Master de traduction,
- Ce premier tiers post-édité sert à adapter le système de TA Anglais-Français. Vu la faible taille des données post-éditées, seuls les poids du modèle log-linéaire sont adaptés en utilisant le premier tiers corrigé comme corpus de développement. Une méthode similaire est suggérée par Pecina et al. (2012) pour l’adaptation au domaine avec quantité de données limitées (nous sommes conscients que d’autres techniques d’adaptation plus avancées auraient pu être utilisées mais ceci n’était pas le thème central de notre contribution),
- Ensuite, le second tiers de l’oeuvre est traduit avec le système de TA adapté, puis la sortie est post-éditée et un second système de TA adapté est obtenu à partir de ces nouvelles données. Ce second système est utilisé pour traduire la troisième et dernière partie de l’oeuvre,
- Une fois la PE terminée, le texte final est révisé : d’abord par le post-éditeur puis par un autre réviseur (non professionnel, francophone, ayant cependant une bonne connaissance de la langue anglaise),
- Les temps de post-édition et de révision sont mesurés.

2.4 Système de TA utilisé

Le système LIG est entraîné à partir des données fournies lors de la campagne d’évaluation IWSLT et optimisé pour la tâche de TA anglais-français de la campagne 2012. Les corpus suivants sont utilisés pour construire le modèle de traduction (total cumulé de 25M phrases environ) : *news-c*, *europarl*, *un*, *ted*, *eu-const*, *dgt-tm*, *pct* et une extraction de 5M de phrases du corpus *gigaword*. La partie en français des mêmes corpus est utilisée pour apprendre le modèle de langue, avec l’ajout du corpus *news-shuffle* fourni au cours de la campagne WMT 2012. Le système LIG est un système à base de segments (*phrase-based*) qui s’appuie sur la boîte à outils Moses Hieu et al. (2007). Trois modèles de traduction appris à partir de différents corpus (*ted*; *news-c+europarl+un+eu-const+dgt-tm+pct*; *gigaword5M*) sont utilisés. Un modèle de langue 5-gramme est appris séparément sur chaque corpus à l’aide de la boîte à outils SRILM Stolcke (2002) avec un modèle de repli Kneser-Ney modifié, puis les modèles sont interpolés en optimisant la perplexité sur le corpus *dev2010* de la campagne IWSLT. Le système du LIG obtient des scores BLEU de 36,88 et 37,58 sur les corpus *tst2011* et *test2012* de IWSLT respectivement (BLEU évalué avec casse et ponctuation). Plus de détails sur ce système (classé honorablement lors de IWSLT 2012) se trouvent dans Besacier et al. (2012). Il est clair que ce système n’est pas optimal pour traduire des textes littéraires et il serait souhaitable, pour de futurs travaux, de rassembler au moins des textes littéraires en français pour adapter le modèle de langue cible (ou même envisager d’avoir accès aux autres livres et traductions de l’auteur).

2.5 Post-édition

Nous utilisons l’interface de post-édition du LIG proposée initialement par Huynh et al. (2008). Celle-ci a été utilisée pour de nombreux projets : traduction d’articles de l’encyclopédie EOLLS, accès multilingue à des dizaines de sites Web. La figure 1 montre l’interface de post-édition en mode avancé qui permet, pour chaque segment source, de charger plusieurs traductions automatiques (issues de Google, Moses⁷, etc.) et de les corriger. Le temps de post-édition est mesuré et l’historique des corrections est stocké dans une base de données. Cet outil est la colonne vertébrale ayant donné lieu au concept d’iMAG proposé par le LIG (*interactive Multilingual Access Gateway*) qui permet la navigation dans un site ou sous-site Web dans plusieurs langues d’accès, avec amélioration incrémentale et contrôlée de la qualité des traductions (voir le lien en note de bas de page pour plus de détails).⁸

3 Données expérimentales

3.1 Corpus et statistiques de post-édition

L’oeuvre, composée de 545 segments et 10731 mots est divisée en trois blocs identiques. La Table 1 résume le nombre de mots des données source et cible (TA ou PE⁹). Sans surprises, un ratio supérieur à 1,2 est observé entre cible française (TA) et source anglaise. On constate cependant que ce ratio tend à diminuer après post-édition de la sortie française.

7. [ici](#), la sortie de notre système était chargée en priorité par rapport à Google

8. aximag.fr/AXiMAG-homePage.html

9. La post-édition utilisée ici est obtenue après chaque itération du processus ; la dernière étape de révision n’est donc pas prise en compte à ce stade.



FIGURE 1 – Interface de post-édition en mode avancé

TABLE 1 – Corpus source, cible traduite et cible corrigée

Itération (nb. seg)	Anglais (nb. mots)	TA Français (nb. mots)	PE Français (nb. mots)
It.1 (184)	3593	4295	4013
It.2 (185)	3729	4593	4202
It.3 (176)	3409	4429	3912
Total (545)	10731	13317	12127

3.2 Performances du système de TA

La Table 2 résume les performances de TA (mesurées avec BLEU) calculées sur le corpus complet avec les systèmes issus de chaque itération. Le temps de post-édition requis pour chaque bloc est également indiqué. Les scores BLEU, qui sont directement comparables, ne montrent pas de véritable amélioration du système. Il semble donc que l’adaptation des poids seuls (qui donne lieu à des améliorations dans Pecina et al. (2012)) soit peu efficace dans notre cas. Le temps de post-édition diminue cependant légèrement à chaque itération (mais là encore, les différences sont faibles et il est difficile de dire si la diminution du temps de PE est dûe à l’adaptation du système de TA ou à une productivité croissante du post-editeur qui s’adapte à la tâche). Au final, le temps total de PE est estimé à 15h environ.

TABLE 2 – Evaluation automatique (BLEU) sur corpus complet et mesure du temps de PE pour chaque bloc de l’itération

Système TA utilisé	score BLEU (corpus complet)	temps PE (bloc it.)
It.1 (non adapté)	38,95	5h37mn
It.2 (tuning sur bloc 1)	37,51	4h45mn
It.3 (tuning sur bloc 1+2)	39,38	4h35mn

La lecture de l’oeuvre traduite à ce stade (après PE) est peu satisfaisante. En effet, la post-édition est faite “segment par segment”, sans vue d’ensemble sur le texte. Il en résulte une manque d’homogénéité très gênant pour un texte littéraire. Pour cette raison, deux révisions du texte traduit sont également effectuées : une par le post-éditeur lui même (4h) et une par l’auteur de cet article (6h). La version finale de l’oeuvre traduite (qui aura donc été obtenue en 15+4+6=25h) sert de base aux évaluations, plus qualitatives, présentées dans la partie suivante. Les différences entre la version post-éditée brute (15h de travail) et la version révisée (25h de travail) ne sont pas analysées ici, mais pourront l’être dans le futur puisque les deux versions sont rendues disponibles sous *github*.

4 Aller au delà d'une tâche de TA adaptée au domaine

4.1 Le point de vue des lecteurs sur la traduction post-éditée

Neuf lecteurs ont accepté de lire l'oeuvre traduite et ont répondu à un questionnaire, toujours ouvert sur *fluidsurveys.com*¹⁴. La version *pdf* de l'essai traduit ainsi que fichier tableur rassemblant les résultats du sondage sont également rendus disponibles dans *github*. Après trois questions permettant de mieux cerner le profil du lecteur, une première partie (5 questions) interroge les lecteurs sur la lisibilité et la qualité du texte littéraire traduit. Une seconde partie (7 questions) vérifie que certaines subtilités du texte ont été bien comprises.

Le texte est jugé globalement lisible (5 TBien et 3 Bien), compréhensible (8 oui, 1 non) et contenant peu de fautes (8 rarement, 1 souvent). **Les questions de compréhension les plus faciles sont bien traitées par les lecteurs qui répondent tous correctement (4 questions).** Cependant, trois questions donnent lieu à des réponses différentes selon les lecteurs :

- 2 lecteurs ont mal répondu à une question apparemment simple (*qui finance le séquençage du génôme de Powers ?*),
- la question *Lors de l'écriture de la nouvelle, combien de personnes se sont déjà fait séquençer le génôme ?* était ambiguë puisqu'on pouvait répondre 8 ou 9 (en comptant ou non Powers) et a donné lieu à des réponses différentes des lecteurs
- seuls 4 lecteurs sur 9 ont su donner une chronologie correcte des étapes du processus de séquençage d'un génôme ; **le texte traduit ne contient cependant pas de contresens ;** ce résultat mitigé indique peut-être un désintérêt de certains lecteurs pour les aspects les plus techniques de l'oeuvre.

4.2 Le point de vue du traducteur de R. Powers

Pour finir cette étude pilote, un dixième lecteur a été sollicité : le traducteur français de l'auteur, J-Y Pellegrin, enseignant chercheur à Paris-Sorbonne. Son avis est résumé ici sous la forme de questions réponses. Le manque de place ne nous permet pas de commenter ces remarques mais nous pensons qu'elles sont assez explicites pour être délivrées en l'état.

Lisibilité ? *Le texte auquel vous êtes parvenu restitue une image fidèle du contenu de l'article de Powers. Le pari de la lisibilité est gagné et certains passages (notamment ceux qui portent sur les aspects scientifiques de l'expérience décrite) sont très convaincants. "*

Imperfections ? *"Il reste bien sûr des imperfections, des lourdeurs, voire des erreurs ponctuelles, qui appellent une correction"*

Principales erreurs ?

- *"Le défaut le plus répétitif, celui dont souffre d'ailleurs le travail de tout traducteur débutant, est le calque syntaxique, là où le français structure différemment la phrase .../... On comprend, mais ça ne sonne pas vraiment français*
- *"Autre défaut assez fréquent, la perte des idiomatismes du français au profit d'anglicismes. Parfois ces anglicismes peuvent être plus déroutants lorsqu'ils flirtent avec le français comme dans *né* connaissances actionnables ž (p. 18) au lieu de *né* connaissances pratiques / utilisables ž. "*
- *"Un troisième défaut tient à la non prise en compte de certains repères culturels .../... Par exemple, Powers fait plusieurs références à la topographie de Boston qui donnent lieu à des inexactitudes dans la traduction : *né* la rivière Charles ž par exemple (p. 12) qui n'est pas une rivière mais plutôt un fleuve ; c'est pourquoi on traduira par *né* la Charles River ž ou simplement *né* la Charles ž "*

Ce texte pourrait-il servir de base de départ à un traducteur littéraire professionnel ? " *Instinctivement, je serais tenté de répondre non pour l'instant, parce que, dès son premier jet, le traducteur possède des réflexes qui lui permettent de produire un texte plus *né* propre ž que celui auquel vous êtes parvenu .../... Cependant, ce traducteur passera plus de 25 heures à produire les 42 feuillets de 1500 signes correspondant au texte de Powers. À raison de 7 feuillets par jour en moyenne, il faut 6 journées de 8h pour venir à bout du texte .../... Si, en revanche, je pouvais ne travailler que sur votre texte, en oubliant complètement celui de Powers parce que j'aurais la garantie que votre traduction ne comporte aucune erreur, ni oubli, ni aplatissement par rapport à l'original, mais qu'elle demande simplement à être améliorée, rendue plus fluide, dans un français plus authentique, les choses seraient différentes et le gain de temps sans doute considérable. "*

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

5 Conclusion

5.1 Données rassemblées et mise en ligne

Les données de cet article sont disponibles sur le lien suivant github.com/powersmachinetranslation/DATA. On y retrouve notamment :

- les 545 segments source, cible (TA et PE) mentionnés table 1,
- l’oeuvre traduite et révisée en français, ayant été lue par un panel de 9 lecteurs,
- les résultats du questionnaire (9 lecteurs) compilés dans un tableur.

5.2 Commentaires et questions ouvertes

5.2 Commentaires et questions ouvertes Nous avons présenté une première expérience de traduction automatisée d’une oeuvre littéraire (essai en anglais d’une vingtaine de pages). Les résultats issus d’un pipeline TA + PE ont été présentés et, pour aller au delà, les avis d’un panel de lecteurs et d’un traducteur ont été sollicités. Le texte traduit, obtenu après 25h de travail humain, est jugé acceptable par les lecteurs mais l’avis du traducteur professionnel reste mitigé. Cette approche suggère une méthodologie de traduction rapide et “low cost”, analogue aux traductions de sous-titres de séries TV trouvées sur le Web. Pour l’auteur, c’est la possibilité d’avoir son oeuvre traduite dans un plus grand nombre de langues (plusieurs dizaines au lieu d’une poignée - cet essai de R. Powers a d’ailleurs aussi été traduit en roumain avec la même méthodologie). Mais celui-ci est-il prêt à sacrifier la qualité de traduction (et son contrôle sur celle-ci) au prix d’une diffusion plus large de ses oeuvres ?

Pour le lecteur qui ne peut lire l’auteur en langue source, c’est la possibilité d’avoir accès plus rapidement à une traduction (certe imparfaite) de son auteur favori. Pour le lecteur non natif mais capable de lire l’oeuvre en langue source, c’est la possibilité d’avoir une aide sur les parties qu’il a du mal à comprendre. Une dernière chose : le titre de l’oeuvre *The Book of Me* est resté inchangé sur la version française car aucune traduction satisfaisante n’a été trouvée pour illustrer le fait que *book* fait référence ici à un livre mais aussi à l’ADN de l’auteur ; ce paradoxe illustre bien toute la difficulté de la traduction d’une oeuvre littéraire.

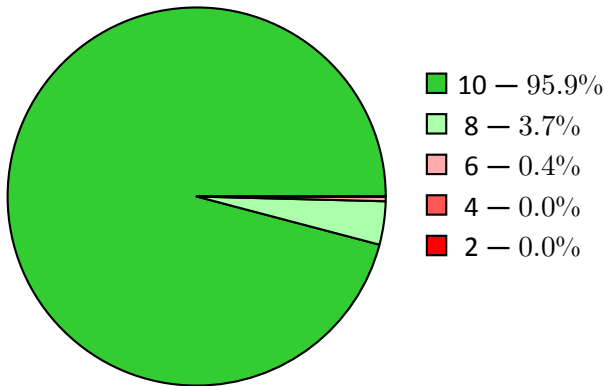
Remerciements

Merci à Manuela Barcan qui a assuré la première phase de post-édition des traductions automatiques en français et en roumain au cours de l’été 2013. Merci à J-Y Pellegrin, traducteur français de Richard Powers, pour son aide et son ouverture d’esprit. Merci à V. Bellynck et C. Boitet pour leur aide avec l’outil de post-édition Sectra_W.

Références

- BESANCIER L., LECOUTEUX B., AZOUZI M. & LUONG NGOC Q. (2012). The LIG English to French Machine Translation System for IWSLT 2012. In *In proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT)*.
- GARCIA I. (2011). Translating by post-editing : is it the way forward ? *Journal of Machine Translation*, **25**(3), 217–237.
- HIEU H., BIRCH A., CALLISON-BURCH C., FEDERICO M., BERTOLDI N., COWAN B., SHEN W., MORAN C., ZENS R., DYER C., BOJAR O., CONSTANTIN A. & HERBST E. (2007). Moses : Open source toolkit for statistical machine translation. In *ACL’07, Annual Meeting of the Association for Computational Linguistics*, p. 177–180, Prague, Czech Republic.
- HUYNH C.-P., BOITET C. & BLANCHON H. (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*, Marrakech, Morocco.
- PECINA P., TORAL A. & VAN GENABITH J. (2012). Simple and effective parameter tuning for domain adaptation of statistical machine translation. In *Proceedings of the 24th International Conference on Computational Linguistics*, p. 2209–2224, Mumbai, India.
- POTET M., EMMANUELLE E.-R., BESACIER L. & BLANCHON H. (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.
- SPECIA L., CANCEDDA N. & DYMETMAN M. (2010). A dataset for assessing machine translation evaluation metrics. In *7th Conference on International Language Resources and Evaluation (LREC-2010)*, p. 3375–3378, Valletta, Malta.
- STOLCKE A. (2002). Srilmm : An extensible language modeling toolkit. In *ICSLP’02, 7th International Conference on Spoken Language Processing*, p. 901–904, Denver, USA.

Translation Adequacy Results



Percentage of 241 evaluated segments judged to be in each translation adequacy category by a bilingual judge.

Automated translation of a literary work : a pilot study

Laurent Besacier ¹

(1) LIG, University of Grenoble, UJF - BP 53, 38041 Grenoble Cedex 9
laurent.besacier@imag.fr

Résumé. Les techniques actuelles de traduction automatique (TA) permettent de produire des traductions dont la qualité ne cesse de croître. Dans des domaines spécifiques, la post-édition (PE) de traductions automatiques permet, par ailleurs, d'obtenir des traductions de qualité relativement rapidement. Mais un tel pipeline (TA + PE) est-il envisageable pour traduire une œuvre littéraire ? Cet article propose une ébauche de réponse à cette question. Un essai de l'auteur américain Richard Powers, encore non disponible en français, est traduit automatiquement puis post-édité et révisé par des traducteurs non-professionnels. La plateforme de post-édition du LIG utilisée permet de lire et éditer l'œuvre traduite en français continuellement, suggérant (pour le futur) une communauté de lecteurs-réviseurs qui améliorent en continu les traductions de leur auteur favori. En plus de la présentation des résultats d'évaluation expérimentale du pipeline TA + PE (système de TA utilisé, scores automatiques), nous discutons également la qualité de la traduction produite du point de vue d'un panel de lecteurs (ayant lu la traduction en français, puis répondu à une enquête). Enfin, quelques remarques du traducteur français de R. Powers, sollicité à cette occasion, sont présentées à la fin de cet article.

Abstract. **Current machine translation (MT) techniques continue to improve.** In specific fields, post-editing (PE) of automatic translations can produce high-quality translations relatively quickly. But can such a (MT + EP) pipeline be used to translate a literary work ? This paper attempts a preliminary answer to this question. A short story by the American author Richard Powers, which is currently unavailable in French, is automatically translated and then post-edited and revised by non-professional translators. The LIG post-editing platform is used to view and edit the short story, potentially enabling a future community of readers and editors to continuously improve translations of their favorite author. In addition to presenting experimental evaluation results (MT system details and automatic evaluation results) for the MT+PE pipeline, we also discuss the quality of the post-edited translation output from the perspective of a panel of readers (who read the translation in French, and then responded to a survey). Finally, some remarks of the French translator of R. Powers, solicited on this occasion, are presented at the end of this article.

Mots-clés : Mots-clés : traduction automatique, TA, œuvre littéraire, post-édition. .

Keywords : Keywords : machine translation, MT, literature, post-editing. .

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 (posteditor) who correct any translation errors.

Numerous studies have shown the benefits of the combined use of machine translation and manual post-editing (MT+PE) in a broadcast stream. 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 experiment² 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.

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

These results from academia (Garcia (2011)) and industry (Autodesk) regarding translation in specialized areas lead us to ask the following question : 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 work of ten thousand words ? Is the resulting translation acceptable to readers ? What would the official translator think of it ? 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).

The following section presents our general methodology, the choice of work and the two systems used, MT and PE (both created from the LIG). 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 an 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

The short format of this article does not allow for a comprehensive literature review on assisted translation of literary documents, or more generally on the use of NLP in the literary field. However, we note that this is a fairly active research area : a workshop has been devoted to this topic since 2012 : *The ACL Workshop on Computational Linguistics for Literature*³. One can also mention the recent *Seminar on the Translation of Literary Text*, organized in France by the ENS in 2013. However, to our knowledge, this study is the first on the automated translation of a literary work. We should now define what constitutes 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 or short stories. 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 express meanings (explicit or implied).

2.2 Choice of work

The choice of work was been guided by the fact that (a) we had a contact with the French translator of the American author Richard Powers⁴ (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 publishes novels linked with science, 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. We then chose a work not yet translated into French, entitled *The Book of Me*, originally published in the pages of GQ magazine⁵. It is a narrative, told 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 irrational.

2.3 General methodology

The translation of texts using the MT+PE (machine translation and post-editing) workflow has previously been evaluated using LIG on journalistic data. In 2012, 12,000 post-edited segments (equivalent to a book of about 500 pages) were collected by Potet et al. (2012). This corpus, collected through *crowdsourcing* and available online⁶, is one of the largest existing corpora 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 the 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.

3. sites.google.com/site/clfl2014a/

4. fr.wikipedia.org/wiki/Richard_Powers

5. www.gq.com/news-politics/big-issues/200810/richard-powers-genome-sequence

6. www-clips.imag.fr/geod/User/marion.potet/index.php?page=download

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

- After automatic translation of the first third (by an initial probabilistic MT system not adapted to this task), the translated output is reviewed by a translation studies Masters student.
- The post-edited first third is used to adapt the English-French MT system. Given the small amount of post-edited data, the weights of the log-linear model are adapted by using the corrected first third as a development corpus. A similar method is suggested by Pecina and 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 work is translated with the adapted MT system, then the results are post-edited and a second adapted MT system is obtained starting from the new data. This second system is used to translate the third and last part of the work,
- Once the PE is 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 are measured.

2.4 MT system used

The LIG 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. The LIG system is a system based on segments (*phrase-based*) which builds on the *Moses* toolkit of Hieu et al. (2007). 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; these models are then interpolated by optimizing perplexity on the IWSLT *dev2010* corpus. The LIG 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) are in Besacier et al. (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 other works and translations by the author).

2.5 Post-editing

We use the LIG post-editing interface initially proposed by Huynh et al. (2008). It has been used for many projects : translation of from the EOLLS encyclopedia, multilingual access to dozens of websites. Figure 1 shows the post-editing interface in advanced mode that allows, for each source segment, multiple automatic translations (from Google, Moses⁷, etc..) to be loaded and corrected. The 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 (*interactive Multilingual Access Gateway*) concept proposed by the LIG which allows browsing a site or subsite in several access languages, with incremental improvement and quality control of the translations (see link at footnote on page for more details).⁸

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 PE⁹) words in the data. Not surprisingly, a ratio greater than 1.2 is observed between French

7. here, the output of our system is loaded in preference to that of Google

8. aximag.fr/AXiMAG-homePage.html

9. 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.



FIGURE 1 – Post-editing interface in advanced mode

TABLE 1 – Source corpus, translated target, and corrected target

Iteration (no. seg)	English (no. words)	French TA (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

target (MT) and English source words. Notably, 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) 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, 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 increasing productivity as the post-editor adapts to the task). In the end, the total PE time is estimated at about 15 hours.

TABLE 2 – Automatic Evaluation (BLEU) on full corpus, along with time measurements for each PE block iteration

MT system used	BLEU score (full corpus)	PE (block it.) time
It.1 (not adapted)	38.95	5h37mn
It.2 (tuning on Block 1)	37.51	4h45mn
It.3 (tuning on Blocks 1+2)	39.38	4h35mn

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) are not analyzed here, but such analysis will be possible in the future since both versions are made available on *github*.

4 Going beyond the task of MT domain adaptation

4.1 The views of readers on the post-edited translation

Nine readers agreed to read translated work and answered a questionnaire, available on *fluidsurveys.com*¹⁴. 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, the first portion (5 questions) asks readers about readability and quality of the translated literary text. The second portion (7 questions) verifies that certain subtleties of the text were understood.

The text is considered to be overall readable (5 Very Good and 3 Good), comprehensible (8 yes, 1 not) and containing few faults (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 of the work aspects.

4.2 The views of R. Powers’s 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. Lack of space does not allow us to comment on these remarks but we think that they are sufficiently clear to be delivered in the state.

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

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, such as translating *řactionable knowledgež* as *řconnaissances actionnablesž* (p. 18) instead of *řconnaissances pratiques / utilisables.ž*"*
- *"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 *řune rivièrèž* but *řun fleuvež* ; that is why we translate *řla Charles Riverž* or simply *řla Charlesž*"*

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 Powell’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) 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 considerable."*

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

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 and PE) segments mentioned in Table 1
- The translated and revised work, in French, that was read by a panel of 9 readers
- The results of the questionnaire (9 readers) compiled in a spreadsheet

5.2 Comments and open questions

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 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 represent 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 V. Bellyncck and C. Boitet for their assistance with the Sectra_W post-editing tool.

Références

- BESANCIER L., LECOUTEUX B., AZOUZI M. & LUONG NGOC Q. (2012). The LIG English to French Machine Translation System for IWSLT 2012. In *In proceedings of the 9th International Workshop on Spoken Language Translation (IWSLT)*.
- GARCIA I. (2011). Translating by post-editing : is it the way forward ? *Journal of Machine Translation*, **25**(3), 217–237.
- HIEU H., BIRCH A., CALLISON-BURCH C., FEDERICO M., BERTOLDI N., COWAN B., SHEN W., MORAN C., ZENS R., DYER C., BOJAR O., CONSTANTIN A. & HERBST E. (2007). Moses : Open source toolkit for statistical machine translation. In *ACL'07, Annual Meeting of the Association for Computational Linguistics*, p. 177–180, Prague, Czech Republic.
- HUYNH C.-P., BOITET C. & BLANCHON H. (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*, Marrakech, Morocco.
- PECINA P., TORAL A. & VAN GENABITH J. (2012). Simple and effective parameter tuning for domain adaptation of statistical machine translation. In *Proceedings of the 24th International Conference on Computational Linguistics*, p. 2209–2224, Mumbai, India.
- POTET M., EMMANUELLE E.-R., BESACIER L. & BLANCHON H. (2012). Collection of a large database of french-english smt output corrections. In *Proceedings of the eight international conference on Language Resources and Evaluation (LREC)*, Istanbul, Turkey.
- SPECIA L., CANCEDDA N. & DYMETMAN M. (2010). A dataset for assessing machine translation evaluation metrics. In *7th Conference on International Language Resources and Evaluation (LREC-2010)*, p. 3375–3378, Valletta, Malta.
- STOLCKE A. (2002). Srilmm : An extensible language modeling toolkit. In *ICSLP'02, 7th International Conference on Spoken Language Processing*, p. 901–904, Denver, USA.

- Joshua S. Albrecht, Rebecca Hwa, and G. Elisabeta Marai. Correcting automatic translations through collaborations between MT and monolingual target-language users. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, pages 60–68, Athens, Greece, March–April 2009.
- Chris Callison-Burch. Linear B system description for the 2005 NIST MT evaluation exercise. In *Proceedings of the NIST 2005 Machine Translation Evaluation Workshop*, 2005.
- Philipp Koehn. Enabling monolingual translators: Post-editing vs. options. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 537–545, Los Angeles, California, June 2010.
- Linda Mitchell, Johann Roturier, and Sharon O’Brien. Community-based post-editing of machine translation content: monolingual vs. bilingual. In *Proceedings of the 2nd Workshop on Post-editing Technology and Practice (WPTP-2)*, pages 35–43, Nice, France, September 2013. EAMT.
- Lane Schwartz, Timothy Anderson, Jeremy Gwinnup, and Katherine M. Young. Machine translation and monolingual postediting: The AFRL WMT-14 system. In *Proceedings of the Ninth Workshop on Statistical Machine Translation*, pages 186–194, Baltimore, Maryland, June 2014. Association for Computational Linguistics.