Post-editing of Technical Terms based on Bilingual Example Sentences

Elsie K. Y. Chan elsie.chan@mail.com
John S. Y. Lee jsylee@cityu.edu.hk
Department of Linguistics and Translation, City University of Hong Kong

Chester Cheng chester.cheng@gmail.com
Department of Translation, The Chinese University of Hong Kong

Benjamin K. Tsou rlbtsou@cityu.edu.hk
Department of Linguistics and Translation, City University of Hong Kong
Chilin (HK), Ltd.

Abstract

As technical fields become ever more specialized, and with continuous emergence of novel technical terms, it may not be always possible to avail of bilingual experts in the field to perform translation. This paper investigates the performance of bilingual non-experts in Computer-Assisted Translation. The translators were asked to identify and correct errors in MT output of technical terms in patent materials, aided only by example bilingual sentences. Targeting English-to-Chinese translation, we automatically extract the example sentences from a bilingual corpus of English and Chinese patents. We identify the most frequent translation candidates of a term, and then select the most relevant example sentences for each candidate according to semantic similarity. Even when given only two example sentences for each translation candidate, the non-expert translators were able to post-edit effectively, correcting 67.2% of the MT errors while mistakenly revising correct MT output in only 17% of the cases.

1 Introduction

Post-editing of machine translation (MT) system output is now commonly incorporated as part of the workflow in the translation industry, since it can produce higher quality texts than manual translation (Garcia, 2011; Green et al., 2013). For texts in scientific or technical domains, it would be ideal to have bilingual domain experts to perform the post-editing. Given the large number of specialized domains and language pairs, however, translators with the required skills are unfortunately not always available. It is therefore important to understand whether those without the full linguistic or technical background could still perform post-editing adequately. While previous research has explored the feasibility of monolingual post-editing (Mitchell et al., 2013), few studies have investigated how well bilingual non-experts can post-edit MT output of technical texts.

This paper evaluates the performance of bilingual novice translators in identifying and correcting MT errors in technical term translation. To simulate a realistic scenario with time constraints, the translators are aided with only a small number of bilingual example sentences from a database of patents. These example sentences are automatically retrieved from PatentLex, a

Proceedings of Machine Translation Summit XIX, Vol. 1: Research Track, pages 385–392
September 4–8, 2023, Macau SAR, China.

large corpus of English and Chinese patents (Lu et al., 2009; Tsou et al., 2019), to illustrate the most likely translation candidates. Results show that, despite their unfamiliarity with the domain, the translators managed to correct a majority (67.2%) of the MT errors, and mistakenly revised correct MT output in less than 17% of the cases.

The rest of the paper is organized as follows. After sketching the research background (Section 2), we describe the translation texts and bilingual example sentences (Section 3). We then define the translation task (Section 4) and report the results (Section 5).

2 Research Background

Although MT systems can often provide high-quality output, high-stakes translation assignments still require manual verification and editing. This can be a challenging task, especially in technical translation when the human translator is not an expert in the field. In this case, the translator may need to consult existing bilingual examples in context, in order to evaluate different translation options of a term (Bowker and Barlow, 2008). Previous research has studied how well bilingual concordancing can assist novice translators in post-editing MT output of patents (Lee et al., 2020). An in-domain bilingual corpus was shown to yield better translation quality than a general-domain one, but MT outputs from Google and Baidu outperformed the post-edited versions in terms of both BLEU score and term accuracy. However, these results may not be conclusive because of possible variations in the concordancing process. The post-editing outcome could be significantly affected by the skills of the individual subjects in discerning relevant bilingual examples, and the amount of time and effort invested.

Our study mitigates these confounding factors by controlling the post-editing time and the set of bilingual examples provided (Section 4). We use PatentLex, a very large corpus of over 300K comparable Chinese and English patents registered in separate jurisdictions, curated within a 10 year period (Lu et al., 2009; Tsou et al., 2019). This corpus has served as the dataset in two Chinese-English patent MT competitions, organized by NTCIR in Tokyo in 2009 and 2010, and won second place in the 2019 Game Changer Innovation Contest organized by TAUS in Singapore. Reflecting its high quality, MT models trained on this corpus have been shown to outperform generic MT tools such as Google, Baidu and Microsoft in patent translation.

3 Data

3.1 Translation materials

We selected 12 patents in English from PatentLex as the materials for this experiment. The professionally translated Chinese versions of these 12 patents served as the gold translation. We used *Google Translate* and *Baidu Fanyi* to automatically translate one passage from each patent (see example in Table 1). In each English passage, two technical terms were highlighted for our subjects to attempt translation. One term required post-editing, and the other term did not:

Post-editing (PE) required The MT system gave a Chinese translation that differed from the gold translation and was incorrect, and therefore required post-editing. For example, the word 进入 *jinru* 'access' in the MT outputs in Table 1 should be revised to 接入 *jieru* 'access'.

Post-editing (PE) unnecessary The Chinese translation given by the MT system was the gold translation or an acceptable alternative, and therefore did not require post-editing. For example, no change was required for the word 组织 *zuzhi* 'tissue' in the MT outputs in Table 1, since it was the gold translation.

Description	Text	Post-editing required	Post-editing unnecessary
Source	Furthermore, by placement of the cuff below the ***access*** site, the fluid collected above the cuff balloon can expose the ***tissue*** on the ***access*** site.	n/a	n/a
MT output (Google)	此外，通过将套囊放置在进入部位下方，收集在套囊气球上方的流体可使暴露部位上的组织暴露。	进入 *jinru* 'access'	组织 *zuzhi* 'tissue'
MT output (Baidu)	此外，通过将袖带放置在进入部位下方，袖带气囊上方收集的液体可以暴露进入部位的组织。	进入 *jinru* 'access'	组织 *zuzhi* 'tissue'
Gold translation (PatentLex)	此外，通过将囊套放置在接入部位下方，囊套气球上所收集的流体可使接入部位上的组织暴露。	接入 *jieru* 'access'	组织 *zuzhi* 'tissue'

Table 1: Excerpt from a passage in the original English patent, its human (gold) translation in Chinese from PatentLex, and the MT output from Google and Baidu. For each highlighted term in the English passage ('access' and 'tissue'), the subjects were asked to decide whether and how to post-edit the MT translation (*jinru* and *zuzhi*), based on bilingual examples (Table 2). The gold translation was not provided to the subjects.

3.2 Bilingual examples

Since the subjects were non-experts, they needed to examine bilingual example sentences to determine whether post-editing was needed. To support the subjects in making well-informed decisions, these examples should include the most likely translation candidates and illustrate the typical context in which the candidate could be used. For each PE-required word and each PE-unnecessary word, we used LexiScan (Tsou et al., 2019) to find the most frequent Chinese renditions in the database. For each rendition, we retrieved all bilingual example pairs and ranked them according to cosine similarity with the source sentence.

In principle, if there were no time constraint, the more examples are viewed by translators, the higher the post-editing accuracy could be expected. In practice, however, translators are under time pressure to deliver their assignments quickly. To simulate a realistic scenario, we provided only 12 bilingual example pairs for each term, comprising 2 examples for each of the 6 most frequent Chinese renditions. Table 2 shows the 6 renditions of the highlighted term 'access' in the example passage in Table 1, as well as one of the two bilingual example pairs provided to the subjects for each rendition.

4 Experimental set-up

4.1 Subjects

Our study involved 61 students enrolled in a Master of Arts programme in translation studies in Hong Kong. Most of them newly or recently completed their undergraduate studies with a non-science major, and were therefore unlikely to be familiar with the subject domain of the translation materials (Section 3.1).

The students were divided into two groups of 20 students each and one group of 21 students. Each student was asked to complete a translation task: answer a distinctive set of 4 translation questions designed in the same format, without the use of any dictionary or reference sources other than the MT outputs (Section 3.1) and bilingual examples (Section 3.2)

Translation candidate	Bilingual example pairs
访问 *fangwen* 'access'	The bioreactor further includes a second substrate, wherein the second substrate is positioned adjacent to the first surface of the first substrate and defines a plurality of connection channels, each of the connection channels being formed so as to be in fluid communication with a corresponding one of the inlet port, the outlet port, the auxiliary port, and the ***access*** port.
	所述生物反应器进一步包括一第二基片，其中所述第二基片位于邻近所述第一基片的所述第一表面处，并界定多个连接通道，每个所述连接通道的形成以使所述输入口、所述输出口、所述辅助口和所述访问口中相应的一个进行液体传送为宜。
接入 *jieru* 'access' (Gold)	It would be desirable to provide a laparoscopic ***access*** apparatus that would maintain a seal against the escape of gas from within a body cavity, that would enable large tissue samples to be withdrawn through the catheter without damage to the pressure seal, and that would also adapt to a variety of instrument sizes and configurations that are to be passed into and out of the catheter.
	因此需要提供一种剖腹接入装置，该装置能够维持密封防止气体从体腔内逸出，能够使大的组织取样通过导管取出而不损坏压力密封，还能够适合进出导管的多种器械尺寸和结构。
存取 *cunqu* 'access'	With a chosen area of the bottle designed to be flexible, a membrane switch, or any other type of pressure sensor, can be fitted to respond to the change of internal pressure within the bottle, when the ***access*** seal is broken, thus providing a method of interfacing the action of opening the bottle with a circuit.
	由于所选择的瓶区域设计为柔性的，所以膜片开关或任何其它类型的压力传感器可被固定，以在打破存取密封时相应于瓶的内部压力变化，从而提供使开启瓶的动作与电路连接的方法。
进入 *jinru* 'access'	The percutaneous ***access*** sheath may be used in conjunction with a deployment catheter, which is provided with a balloon at its distal end.
	可以与在其远端设置有气囊的扩展导管相结合地使用经皮进入套管。
入口 *rukou* 'access'	As disclosed herein, the ribbon holder includes a cover to allow ***access*** to the through passage whereby the ribbons can be placed into the passage transversely thereof.
	如这里所揭示的，所述光缆支架包括一个压盖，可形成进入通道内的入口来置入所述光缆的横截面。
接近 *jiejin* 'access'	The clamp is structured to contact the diaphragm along a perimeter portion and allow ***access*** to a center portion of the diaphragm.
	该夹具沿周边部分接触振动膜，并且允许接近振动膜的中心部分。

Table 2: Bilingual example pairs for the word 'access' in the passage in Table 1, intended to illustrate the usage context of the top six translation candidates *fangwen*, *jieru*, *cunqu*, *jinru*, *rukou* and *jiejin*

provided on the question paper.

4.2 Translation task

The study was conducted on paper, in the form of 3 typed question papers (namely, Post-editing Exercises A, B and C) each containing 4 translation questions. The 12 distinctive translation questions (namely, Sentences A1-A4; B1-B4; C1-C4) correspond to different text segments taken from 12 selected patents covering varied technical domains.

Regarding the 4 questions in each of the 3 question papers, each question contains: (1) a distinctive patent excerpt in English taken from the corpus of PatentLex, which contains two specific words being highlighted; (2) the corresponding automatic MT translations in Chinese by *Google Translate* and *Baidu Fanyi*, to be used as the reference for the translation task; and (3) for each highlighted word, 12 bilingual English-Chinese sentences taken from other patents from PatentLex that contain the same English word, to be used as the reference for their translation task. For the two highlighted words in each patent excerpt, one is PE-required and the other PE-unnecessary, a fact which was not made known to the subjects. The 12 questions altogether feature 24 highlighted English words to be translated into Chinese with the MT translations and PatentLex bilingual sentences as the only reference sources.

The subjects were asked to work independently on their own translation task, which features texts from filed patents on technical domains that are likely to be unfamiliar to them. After a brief introduction by the instructor, the subjects were given 30 minutes to determine 'the most appropriate translation' of the highlighted English words in Chinese, and another 5 minutes to input their answers on a designated Google Form.

4.3 Manual assessment

The subjects' translations gathered via the designated Google Form were reviewed by two human judges, both native speakers of Chinese. One judge was an experienced translation teacher and professional translator with a PhD in translation studies, who administered the experiment. The other judge was a PhD candidate in Translation with considerable experience working with translation of English-Chinese technical texts.

The judges considered all the translations provided by the subjects and on average accepted 1.5 alternative translations (ranging from 0 to 3) for each highlighted English word, in addition to the gold translation presumably provided by professionals. The judges reconciled the final decision through discussion.

5 Results

Our post-editing study involved 488 instances of term translation from English to Chinese. These included 244 instances that required post-editing (PE) (left side of Table 3) and 244 instances for which post-editing was unnecessary (right side of Table 3). Overall, the subjects achieved 75.4% accuracy by correctly revising 164 of the 244 PE-required instances, and correctly keeping the MT output in 204 of the 244 PE-unnecessary instances. The quality of the post-edited translation was thus higher than the MT output (without post-editing), which had 50% accuracy among the 488 instances. This result shows that, even without the provision of necessary contextual information of the patent excerpts, the subjects were fairly able to deduce the meaning of the highlighted words from the MT translations and bilingual sentences provided and consequently infer either the gold translations or acceptable translations.

5.1 PE-required cases

As shown in Table 3, out of the 244 PE-required cases, the subjects correctly post-edited 164 translations, representing an average accuracy rate of 67.2%. The result is fairly satisfactory,

Text	Post-editing required				Post-editing unnecessary			
	Term	Cor.	Inc.	Cor. %	Term	Cor.	Inc.	Cor. %
A1	access	20	0	100.0%	tissue	12	8	60.0%
A2	reservoir	18	2	90.0%	simulation	19	1	95.0%
A3	access	17	3	85.0%	data	19	1	95.0%
A4	reservoir	4	16	20.0%	body	16	4	80.0%
B1	configuration	10	11	47.6%	slot	13	8	61.9%
B2	operation	2	19	9.5%	system	14	7	66.7%
B3	function	17	4	81.0%	module	19	2	90.5%
B4	act	15	6	71.4%	surface	20	1	95.2%
C1	control	12	8	60.0%	barrier	20	0	100.0%
C2	position	17	3	85.0%	transaction	18	2	90.0%
C3	amount	18	2	90.0%	functional group	16	4	80.0%
C4	switch	14	6	70.0%	solution	18	2	90.0%
Total		164	80	67.2%		204	40	83.6%

Table 3: Post-editing results: number of correct (cor.) and incorrect (inc.) instances among post-editing required terms and post-editing unnecessary terms

given the fact that most subjects did not possess a technical or science background.

Falling slightly below the said average are the translations for 'control' (C1), with a passing 60% (the gold translation being 对照 *duizhao* as in 对照器件 *duizhao qijian* 'control devices'; both MT translations being 控制 *kongzhi* as in 控制装置 *kongzhi zhuangzhi*, which is considered marginally acceptable); while the remaining 40% are inaccurate (调节 *diaojie*) and imprecise (对照变量 *duizhao bianliang*) renderings. Faring less well is 'configuration' (B1, with an accuracy rate of 47.6%), which refers to paper notes in folded shape, form, state or arrangement (形状 *xingzhuang* being the gold translation; 配置 *peizhi* - the MT translation - and *zhuangtai* - proposed rendering by 2 subjects - are considered marginally acceptable); the wrong translations attempted (构形, 结构, 模型, 装置) indicate unacceptable collocation with the word 'note' and inadequate comprehension of context.

A more challenging word is 'reservoir' (A4), with mere 20% accuracy. The obvious reason is that in that particular context, 'reservoir' refers to a sample of aqueous body (水体 *shuiti* being the gold translation) which can be as large as 'Umberumberka Reservoir' (水库 *shuiku* being the gold translation) or as small as that on 'laboratory film balances.' Both judges found it difficult to find one Chinese word that collocates with both sample types: both the MT translation (储层 *chuceng*) and the other attempted translations by the subjects (储器, 储罐, 储液器, 储水器, 蓄水池) appear out of place. The judges suggested that different translations be adopted for the specific 'Reservoir' (水库 *shuiku*) and the generic 'reservoir' (贮库 *zhuku* being an acceptable alternative by 3 subjects; or 储体 *chuti* or 采样来源 *caiyang laiyuan* proposed by the judges) for better textual cohesion and consistency.

The worst performance is for 'operation' (B2), with a meagre 9.5% accuracy rate. Although the respective terms 'interventionist operation' and 'endovascular operations' may appear distinguishing, only 2 out of 21 subjects could infer the gold translation (手术 *shoushu*). 10 subjects adopted the MT version (操作 *caozuo*), which is imprecise for a technical patent, while the rest proposed incoherent translations (操作系统, 操刀, 生产, 运行, 作业, 装置, 反应), reflecting inadequate comprehension caused probably by a lack of technical knowledge and vocabulary.

It is worth noting that a good number of the MT translations were considered to be marginally to reasonably acceptable by the two judges, even though they deviate from their

respective gold translations. Examples include those in A2 (储层 *chuceng* vs. 油藏 *youcang* 'reservoir'), A3 (访问 *fangwen* vs. 存取 *cunqu* 'access'), B1 (配置 *peizhi* vs. 形状 *xingzhuang* 'configuration'), B3 (功能 *gongneng* vs. 函数 *hanshu* 'function'), C1 (控制 *kongzhi* vs. 对照 *duizhao* 'control'), C2 (头寸 *toucun* vs. 立场 *lichang* 'position') and C4 (交换 *jiaohuan* vs. 切换 *qiehuan* 'switch'). All in all, the MT translations and PatentLex bilingual sentences provided, albeit decontextualized and containing one supposedly wrong translation that deviates from the gold standard, are found to be fairly helpful to the subjects for completing their post-editing task with a satisfactory accuracy rate under stringent conditions – translating 8 words used in a technical sense taken from 4 decontextualized patent excerpts within 30 minutes without access to dictionaries or translators' normal reference tools, not to mention the fact that most of the subjects did not possess a technical or science background.

5.2 PE-unnecessary cases

Out of the 244 PE-unnecessary cases, the subjects correctly kept the MT versions, which are the same as the gold translation, in 204 cases, representing a high average accuracy rate of 83.6%. This illustrates the overall efficiency of the selected MT and PatentLex texts and the subjects' post-editing capability. However the accuracy scores for 'tissue' (A1, 60%), 'slot' (B1, 61.9%) and 'system' (B2, 66.7%), albeit middling in absolute terms, are relatively lower. Below is an analysis of the translation errors, which points to the significance of contextual understanding and subject knowledge for the translator.

For 'tissue,' the wrongly attempted translations (组织部位, 体组织, 织物, 薄纸) reflect the subjects' failure to comprehend the context or subject matter – tracheal tissue (组织 *zuzhi* in the MT version) being exposed in relation to a medical device. For 'slot' as in 'note entry slot' (the MT version being 槽 *cao* as in 钞票进槽 *chaopiao jincao*), the unacceptable renderings proposed (狭槽, 槽缝, 狭缝, 缝隙, 缝) reflect imprecision or mis-collocation on the part of the subjects.

The term 'system' refers to a 模拟系统 *moni xitong* 'simulation system' for simulating an 'interventional operation' (模拟介入操作 *moni jieru caozuo* in the MT version) and 'endovascular operations' (血管内操作 *xueguan nei caozuo* in the MT version) using a device (装置 *zhuangzhi* and 设备 *shebei* in the MT versions) equipped with the patented invention. The subjects who correctly kept the MT translations of this term understood that it concerns medical operations. However, the wrong translations 装置 *zhuangzhi* and 装备 *zhuangbei* for 'system' show that the subjects concerned failed to notice that the said Chinese versions should be reserved for 'device' in the same sentence, while the other wrong translations 环境 *huanjing* and 方法 *fangfa* indicate the other subjects' inadequate understanding of the context and technical subject.

6 Conclusion

As technical fields become ever more specialized, and with continuous emergence of novel technical terms, it may not be always possible to avail of bilingual experts in the field to perform translation. In the age of artificial intelligence, translators are increasingly expected to function as post-editors. This paper has investigated the performance of bilingual but non-expert translators in post-editing. Targeting English-to-Chinese translation of technical terms in patents, we asked translators to post-edit these terms in MT output, aided only by bilingual example sentences that were automatically extracted from the PatentLex database.

The results show that, even in the absence of dictionaries and field knowledge, the subjects were in general fairly able to deduce word meaning and produce acceptable translations (75.4%) in decontextualized technical texts with the help of MT translations and the bilingual corpus of PatentLex. In particular, they corrected a majority (67.2%) of the MT errors, and mistakenly

revised correct MT output in less than 17% of the cases. The understanding of context and field knowledge remains crucial for highly accurate and professional translation. In the future, we plan to conduct larger-scale experiments to further shed light on the increasing efficiency and reliability of MT translation and bilingual corpora as indispensable tools for translators.

Acknowledgments

We gratefully acknowledge support from the Strategic Research Grant (projects #7005709 and #7005803) at City University of Hong Kong; and from Hong Kong's Innovation and Technology Commission (ITC) grant to Chilin (HK) Ltd for the PaTTA project (ITC/ESS Project: B/E019/20).

References

Bowker, L. and Barlow, M. (2008). A comparative evaluation of bilingual concordancers and translation memory systems. In Rodrigo, E. Y., editor, *Topics in Language Resources for Translation and Localisation*, pages 1–22. John Benjamins, Philadelphia, PA.

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

Green, S., Heer, J., and Manning, C. D. (2013). The Efficacy of Human Post-Editing for Language Translation. In *Proc. CHI*.

Lee, J., Tsou, B., and Cai, T. (2020). Using Bilingual Patents for Translation Training. In *Proc. 28th International Conference on Computational Linguistics (COLING)*.

Lu, B., Tsou, B. K., Zhu, J., Jiang, T., and Kwong, O. Y. (2009). The Construction of a Chinese-English Patent Parallel Corpus. In *Proc. MT Summit XII: Third Workshop on Patent Translation*, pages 17–24.

Mitchell, L., Roturier, J., and O'Brien, S. (2013). Community-based post-editing of machine-translated content: monolingual vs. bilingual. In *Proc. MT Summit XIV Workshop on Post-editing Technology and Practice*.

Tsou, B. K., Chow, K., Nie, J., and Yuan, Y. (2019). From the cultivation of comparable corpora to harvesting from them: A quantitative and qualitative exploration. In *Proc. 12th Workshop on Building and Using Comparable Corpora*.