

# Teaching MT Through Pre-editing: Three Case Studies

Chi-Chiang Shei

Department of Foreign Languages and Literature

National Tsing Hua University

101, Sec.2, Kuang-Fu Road, Hsinchu 30013

Taiwan

ccshei@mx.nthu.edu.tw

## Abstract

This article reports on three cases of teaching translation or English as a foreign language using pre-editing tasks with a machine translation system. Trainee translators or English learners were asked to input a Chinese or English paragraph into an MT system, observe the irregularities in the output, and subsequently edit the source text and input it again in the hope of getting better output from the MT system. Student reports on pre-editing procedures were collected and analysed, as well as thoughts and suggestions about using MT. It is argued that pre-editing MT input can boost student learning in the cognitive and affective domains, apart from training students to use MT systems.

## 1. Introduction

It has been explained elsewhere by the author (Shei, 2002) that translating into a second language (L2), although not a preferred direction of translation, is nevertheless necessary due to the lack of native speakers in local settings where the service is required. In this context learning an L2 is a subset of learning to translate into the L2. Also, translation is increasingly being reconsidered as a viable means for learning an L2 (Cook, 1998). Thus it can be argued that L2 learning and learning to translate become more and more intertwined. This article explores the use of MT not only in the context of translation teaching but also in the context of L2 learning, showing that MT can be “a

useful engine” in both settings, probably providing a convenient link for the two intellectual activities.

Based on the types of input into an MT system, three categories of pre-editing tasks are distinguished in this article.

- Pre-editing text written in mother tongue: Input is a Chinese text written by the native speaker. Output is the English text produced by the MT system. Object is to modify the Chinese text until the English translation is in the best condition.
- Pre-editing L2 text written by native speaker: Input is an English text written by the native speaker. Output is the Chinese text produced by the MT system. Object is to modify the English text until the Chinese translation is in the best condition.
- Pre-editing L2 text written by learner: Input is an English text written by the learner. Output is the Chinese text produced by the MT system. Object is to modify the English text until the Chinese translation is in the best condition.

## 2. Pre-editing mother language

A class session in the General Translation module instructed by the author was devoted to pre-editing MT input. The direction of translation was from Chinese to English. There were two MT systems used (from which the student should choose one) which offered free online translation facility: 1. SYSTRAN at <http://babelfish.altavista.com/translate.dyn>,

(Babel Fish Translation)<sup>1</sup> and 2. Beijing Golden Bridge Translation Port Network at <http://www.netat.net/>. Here we will use a simple paragraph to illustrate the pre-editing procedure taught in this class. In (1) we have a randomly chosen Chinese paragraph from a randomly picked Web page:

(1)

(a) Source text<sup>2</sup>

十四歲許姓少女將歌手孫燕姿當偶像，認為瘦就是美，控制飲食並運動減重，從四十五公斤減到廿八公斤，外型變成皮包骨，有天體力不支昏倒，被送到醫院急救，回到家仍繼續減重。

(b) Target text in English

14 year old of little surname few heroic women singer grandson swallow posture works as the idol, thinks thin is beautiful, the control diet and the movement reduces weight, reduces from 45 kilograms to 28 kilograms, the outlook turns the skin and bones, has day physical strength does not faint, is delivered the hospital first aid, returned proficiently still continues to reduce weight.

As we can see from (1b), the English translation by SYSTRAN based on (1a) has quite a few problems caused by literal translation and mistaken analysis of structure. One of the most conspicuous problems may be the run-on sentence phenomenon and the subject-less sentences which are a peculiar characteristic of

---

<sup>1</sup> SYSTRAN is the more often used one in our class for its faster Internet connection. This is also the default MT system used for the examples in this article.

<sup>2</sup> Both SYSTRAN and Golden Bridge accept Simplified Chinese (used in China) only, but for technical difficulties of dealing with Simplified Chinese in a Traditional Chinese based OS (used in Taiwan), the latter is adopted for presentation in this article, though in the actual MT rendering process, Simplified Chinese was the one used.

Chinese clauses. To rectify this problem, we need to amend the punctuation (i.e. change a Chinese comma to a period where we think an English sentence should be rounded up) and add some subject nouns to (1a). Thus we have the new set of input and output in (2):

(2)

(a) Source text

十四歲許姓少女把歌手孫燕姿當偶像，她認為瘦就是美，少女控制飲食並運動減重，她的體重從四十五公斤減到廿八公斤，她的外型變成皮包骨，少女有一天體力不足昏倒，被送到醫院急救，她回家後仍繼續減重。

(b) Target text in English

14 year old of little surname young girl the singer grandson swallow posture works as the idol, she think thin is America, the young girl controls the diet and the movement reduces weight, her body weight reduces from 45 kilograms to 28 kilograms, her outlook turns the skin and bones, the young girl one day physical strength insufficient faints, is delivered the hospital first aid, after she went home still continue to reduce weight.

If we compare (1b) and (2b), we find (2b) to be generally better in that each sentence is clearly delimited and most are equipped with a proper subject. Some small adjustments to (1b) also produced improved results in (2b). For example, the erroneous ‘few heroic women’ (*shao nv jiang*) due to a wrong detection of word-boundary was replaced by the correct ‘young girl’ etc. when the original character 將 (*jiang*) was changed to its synonym 把 (*ba*)<sup>3</sup>.

Thus the students went through the

---

<sup>3</sup> For some unknown reason, however, the originally better ‘thin is beautiful’ was replaced by the worse ‘thin is America’ (though these are represented by the same characters in Chinese) even though no adjustment has been made in this respect.

process of modifying the input, observing the output, noting down the limitations of MT, thinking about how to modify the input further, and repeating the sequence over and over, until they obtained satisfactory results (or until they felt the MT output can no longer be improved).

The students were asked to hand in a list of MT flaws they observed afterward, which can be roughly classified as follows:

#### **a. Structural limitations**

- MT often deals with long sentences by (erroneous) chunking, translating each chunk first and then combining chunks to form an illogical sentence.
- MT deals with each sentence independently, i.e. cannot relate consecutive sentences in a meaningful way.
- Complicated noun phrases are problematic for MT systems.
- Most sentences produced by MT are incomplete, literally translated and prone to grammatical errors.
- MT often wrongly assigns the first word in the Chinese sentence as the subject of the target sentence.
- The longer the sentence, the more words are omitted, the more unusual or ambiguous words are used, the poorer the results.
- If the conjunctions in Chinese are omitted, the MT system will not detect the relationship between the connected elements.
- MT frequently fails to detect the relationship between the modifier and the modified, and the relative positions of these two often have to be adjusted to achieve better results.

- Chinese sentences consist of “a lot of verbs”, which are difficult for MT systems. For example, MT cannot correctly detect the boundaries between consecutive verbs.

- A Chinese clause often consists of two (short) English sentences, but MT fails to segment these types of clauses and translates them into single peculiar sentences.

#### **b. Lexico-semantic limitations**

- The MT system used is helpless in determining verb tenses, the number of nouns and parts of speech, references of demonstratives, etc.
- MT systems do not really “understand” the meaning of the Chinese sentences, which can be seen from the fact that literal (and often anomalous) translations prevail.
- MT cannot decide correctly which Chinese character is to be translated into an English preposition.
- Many function words in Chinese have more than one meanings, but MT mostly uses only one (most frequently used) meaning for each of these words in all situations.
- Some words which have meanings are mistakenly translated with non-words with similar phonetic representations.

#### **c. Idiomatical limitations**

- MT systems cannot handle most Chinese idioms, metaphors, colloquial expressions, borrowings from other Chinese dialects, or older Chinese variants.

#### **d. Cultural limitations**

- MT needs to improve its ability to select appropriate words for a specific

background.

- MT does not consider “habit of speaking” and is not “humanised” enough.

#### **e. Operational limitations**

- MT is inefficient, as the paragraph given can be well translated by human within fifteen minutes, but it takes four to five hours to make an MT system produce a marginally-acceptable result through repetitive pre-editing.
- Sometimes the Chinese source sentence must be changed to a “distorted” style in order to get a better English translation.

In addition, students also provided some strategies for improving the performance of MT (in terms of Chinese-to-English translation) based on their pre-editing experience. The following types of pre-editing strategies can be identified:

#### **a. Reorganisation**

- In order to get better English translations, long Chinese sentences are best split into short ones, each having its own subject and verb.
- Changing Chinese sentences to “English style” sentences may achieve better translation results.
- As in human translation, we must chunk the Chinese sentence properly in order for the MT to achieve better results.
- Word order in source sentences must be adjusted to be like that in the target language.

#### **b. Simplification**

- Try to simplify Chinese sentence input, by using fewer and easier words

- The Chinese sentences to be translated should be clear and easy to understand, and without any omissions.

#### **c. Addition**

- Like subjects, the omitted objects of verbs in Chinese must be restored in order to generate complete sentences in English translations.
- Some addition of words to make existing words clearer is helpful.
- Time adverbs implicit in the Chinese text often has to be explicit added in order to get the desired counterpart in English translation.

- Where Chinese does not need an article, an article must nevertheless be added to nouns in the source text in order for the English translation to be correct.

#### **d. Replacement**

- Using pronouns in Chinese achieves better results than using proper nouns.
- Changing part of speech of a word can sometimes yield a better result.
- Replacing negatives with positives can gain better result, for example, changing *can not be defeated* to *can win*.
- Replace Chinese idioms with straightforward expressions.

#### **e. Pre-translation**

- When translating Chinese proper names, it is safer to manually translate them into English first before feeding them into the MT system; otherwise they may cause a confusion in meaning analysis<sup>4</sup>.

---

<sup>4</sup> Chinese proper nouns consist of meaningful

- Before submitting the source text, imagine what the translation will be like first and make necessary alterations.

#### **f. Punctuation**

- Proper manipulation of punctuation can help MT achieve correct structural analyses.

Some advantages for trainee translators doing pre-editing exercises are immediately obvious:

#### **a. Technological awareness**

Trainee translators get to know more about the nature of MT and become better prepared for collaborating with MT systems in the future.

#### **b. Translation competence**

From the processes of editing source text, translation learners are in effect reviewing and applying their translational skills while they are rearranging the words or structures of the source text etc to see how these might affect the MT outcome.

#### **c. Linguistic competence**

While scrutinising MT output in L2, trainee translators as L2 learners are actually reflecting on their own knowledge in the target language, possibly looking up dictionaries etc to verify their hypothesis.

#### **d. Affective boost**

Pre-editing the MT input is an engaging activity, as this involves the operation of some magical machinery, so to speak, especially to students from an arts background. Students are at once

---

characters, and they are not distinguished from other kinds of words by mechanical means such as capitalisation.

fascinated by the automatic translation machine and intrigued by the frequently anomalous output.

It is also obvious from the previous analysis that learner feedback from the pre-editing task may benefit the MT industry and research circles by pointing to areas for improvement.

### **3. Pre-editing Foreign Language**

A part of an English learning module, called Advanced English: English and the Computer, taught by the author to Taiwanese students, concentrates on the use of MT for language learning, among other computer applications. In one of the sessions concerning MT, students were asked to continuously pre-edit an existing English paragraph written by a native speaker in order to get the best MT output. The point of the exercise is two fold: to let the learners grow a feel of how MT can be of help to their L2 learning, and to let them experiment with English words and structures, replacing pieces of native speaker writings with learners' interlanguage while evaluating their hypotheses against the MT output. The report to be handed in consisted of a series of subsequently altered text and their translations, and a list of suggestions as to how to make the MT output more error-free. The following is an example shown to the students:

- (3)  
(a) Source text (English)

What I wish for all students is some release from the clammy grip of the future. I wish them a chance to savour each segment of their education as an experience in itself and not as a grim preparation for the next step. I wish them the right to experiment, to trip and fall, to learn that defeat is as instructive as victory and is not the end of the world. (Lebauer & Scarcella, 1993, p. 91)

(b) Target text (Chinese)

什麼我祝願所有學生是某一發行從未來的陰濕粘粘夾子。我祝願他們機會對滋味他們的教育的各段作為經驗本質上和不是作為一種冷面準備下步。我祝願他們權利試驗，絆倒和落，獲悉，失敗象勝利是一樣教育和不是這個世界的末端。

As the MT output in (3b) has some problems caused by direct translation (in terms of structure) and literal translation (in terms of diction), two immediate ways for improving the translation are to change word order or to restructure the sentence, and to replace troublesome words or phrases with synonyms or similar expressions. For instance, the first sentence in (3a), *What...is...*, is directly translated by MT into some sort of rhetoric question in Chinese in (3b), which is incorrect. So it should be changed into a more straightforward sentence. Also in the first sentence, *release* is translated to the meaning of “publishing” rather than “freeing”, and *grip* was translated with the meaning of “hook” rather than “grasp”. Nominalisation is rare in Chinese, so the action related to “release” is returned to its passive verb form. Thus the first sentence in (3a) can be changed into something like: *I wish all students are released from the clammy grasp of the future.*

(4)

(a) Source text (English)

I wish all students are released from the clammy grasp of the future.

(b) Target text (Chinese)

我祝願所有學生從未來的陰濕粘粘掌握釋放。

And indeed the MT output in (4b) is quite an acceptable Chinese sentence, much better than the first sentence in (3b).

For this exercise, the article taken from Lebauer & Scarcella (1993) had been gone through in detail in class beforehand, from which students were asked to select a paragraph as the input to MT. Since students already had a good understanding of the text chosen, the process of word substitution and restructuring means a kind of “expanded learning” for them, where the newly acquired knowledge in English can be more readily incorporated into their existing knowledge of English. The following is part of a student’s ideas in the reports they turned in:

This work is very interesting and inspiring. I can train my skill of composition and try to know how to rewrite a complicated sentence into a simplified one. Although there might be something wrong in my corrections, but it is still a good experience for me, especially, I can enrich the knowledge about using right words in an article.

This was what the author had expected the students to find. However, there were also less positive feedbacks. For example, one student wrote that the only thing he learned from the exercise was to realize that we could not rely on translation when learning English. Another one said that he found his grammatical competence was as poor as the MT system from doing the exercise.

While in the previous pre-editing task the input was the Chinese (source) text, in the current task the input was the English text, both written by native speakers respectively. In both cases, there was a definite reference point for the students to base their investigations on (i.e. the texts written by native speakers). In the third case to follow, however, the pre-editing task was more demanding, in that the input was the immature learner composition, and the output was the imperfect MT translation. Modifying the source text in order to get better MT output became more difficult since there was no longer a

standard text to work against.

#### 4. Pre-editing learner L2 composition

A session of the General English (GE) module taught by the author to Taiwanese students was devoted to the use of MT to help with student composition in L2 (English). Student first make out their composition in English, input their L2 composition into the MT system, scrutinise the result of translation while attempting to find any irregularities which could have been caused by the errors in their original composition, go back to edit their English composition, and then go through the procedure again and again, until the MT output is acceptable or until there seems no possibility for improvement.

MT in this case is used as a “grammar checker”, resembling a common grammar-checking program developed for learners in the following ways:

- It deals with learner text.
- It flags possible errors in the text.

Yet an MT system used as a grammar checker differs from a proper grammar-checking program in the following ways:

- It offers diagnosis and suggestions in an implicit way.
- Erroneous analyses are intermingled with correct analyses.
- The result is more prone to mistaken interpretation.

The following is one of the twenty learners' compositions, whose name is Jane:

(5)

There are many pressures in a freshman year. My pressure is peer pressure. In the first, I get a good friend in my class, we have the same

hobby and personality. Gradually, I am afraid of walking or talking with her. Because she is very smart, and she make a lot of friends in our class. I often need her helps in mathematic and computer. Whenever I stay with her, I feel I am nothing in the world.

The Chinese translation of (5) offered by SYSTRAN is:

(6)

有許多壓力在 freshman year。我的壓力是同輩壓力。在一，我得到一個好朋友在我的類，我們有同樣愛好和個性。逐漸，我害怕行走或談話與她。由於她是非常聰明，和她做很多朋友在我們的類。我經常需要她幫助在數學和電腦裏。每當我留下與她，我感到我是無事在世界上。

After examining the first MT version, the student made changes and produced the following English text:

(7)

There are many pressures in a freshman year. My pressure is peer pressure. At first, I get a good friend in my class, we have the same hobby and personality. Gradually, I am afraid of walking or talking with her. Because she is very smart, and she make a lot of friends in our class. I often need her helps in mathematic and computer. Whenever I stay with her, I feel I am nothing in the world.

The most significant change in this example is the replacement of *in the first* in (5) with *at first* in (7). Presumably, the student, Jane, found the Chinese translation by MT of *in the first* to be unacceptable – It is translated literally into the nonsensical Chinese *zai-i* (“at one”). She probably started recalling the possible English combination of words she had learned that expressed the same concept, and found *at first* to be a likely option. A later consultation with Jane showed this to be the case. As expected, the Chinese translation produced by MT in (8) contained a proper Chinese word *qichu*

("initially") in contrast to the previous nonsensical *zai-i*.

(8)

有許多壓力在一新生年。我的壓力是同輩壓力。起初，我得到一個好朋友在我的類，我們有同樣愛好和個性。逐漸，我害怕行走或談話與她。由於她是非常聰明，和她做很多朋友在我們的類。我經常需要她幫助在數學和電腦裏。每當我留下與她，我感到我是無事在世界上。

The Chinese version in (8), however, still contains several irregularities. For example, the *in* in *helps in mathematic* was literally and mistakenly translated as *li* ("inside"). Jane noticed this, and, in an effort to accomplish the task assigned by the teacher – i.e. changing the English *in* order to get the best Chinese translation, she replaced the correct preposition *in* with the incorrect *on*. Ironically, this produced a better Chinese translation, as in (10), since the literal *li* was gone.

(9)

There are many pressures in a freshman year. My pressure is peer pressure. At first, I get a good friend in my class, we have the same hobby and personality. Gradually, I am afraid of walking or talking with her. Because she is very smart, she makes a lot of friends in our class. I often need her helps on mathematic and computer. Whenever I stay with her, I feel I am nothing in the world.

(10)

有許多壓力在一新生年。我的壓力是同輩壓力。起初，我得到一個好朋友在我的類，我們有同樣愛好和個性。逐漸，我害怕行走或談話與她。由於她非常聰明，她做很多朋友在我們的類。我經常需要她幫助在數學和電腦。每當我留下與她，我感到我是無事在世界上。

From observing the exercises handed in by the twenty students, some types of pre-editing behaviours can be identified regarding learner L2 composition:

### a. Fixing mechanical problems:

#### Example 1. (Jane)

Original sentence: There are many pressures in a freshman year.

MT translation: 有許多壓力在 freshman year

Revised sentence: There are many pressures in a freshman year.

MT translation: 有許多壓力在一新生年

Reason: *Freshman year* was found to be un-translated.

#### Example 2. (Jennifer)

Original sentence: I don't know whether I could do it well.

MT translation: i don.t 知道是否我能做它很好

Revised sentence: I do not know whether I could do it well.

MT translation: 我不懂得是否我能做它很好

Reason: A non-English punctuation was found to cause non-translation.

### b. Replacing words or phrases:

#### Example 1. (Clark)

Original sentence: I have not had any serious pressure.

MT translation: 我沒有任何嚴肅的壓力 ("I don't have any solemn pressure")

Revised sentence: I have not had any acute pressure.



MT translation: 我沒有任何深刻壓力 (“I don't have any profound pressure”)

Reason: *Serious* was first translated into something unintended.

*Example 2. (Avirl)*

Original sentence: I had found something that made me felt I was under pressure.

MT translation: 我發現了某事做了我毛氈我是在壓力下 (“I found something done I blanket I was under pressure”)

Revised sentence: I had found something that made me feel I was under pressure.

MT translation: 我發現了某事做了我感受我是在壓力下 (“I found something done I felt I was under pressure”)

Reason: *Felt* was found to be mistakenly used due to the anomalous translation.

**c. Adding or deleting words:**

*Example 1. (Avirl)*

Original sentence: I do not live in the dormitory.

MT translation: 我不居住在這棟宿舍

Revised sentence: I do not live in dormitory.

MT translation: 我不居住在宿舍

Reason: *The* was given a demonstrative sense in the Chinese translation, which the learner did not intend.

*Example 2. (Jamie)*

Original sentence: I discover that it is not an easy thing to make a friend.

MT translation: 我發現這不是做一位朋友

的一件容易的事情 (“I found this is not an easy thing for a friend”)

Revised sentence: I discover that it is not an easy thing.

MT translation: 我發現這不是一件容易的事情 (“I found that this is not an easy thing”)

Reason: Original translation yields unintended meaning.

**d. Changing structures:**

*Example 1. (Eva)*

Original sentence: I feel tired by the stress.

MT translation: 我感到疲乏的由重音

Revised sentence: The pressure makes me feel tired.

MT translation: 壓力使我感受疲倦

Reason: *Stress* was translated by the meaning of ‘prominence of sound’ in the first translation.

*Example 2. (Cindy – Golden Bridge)*

Original sentence: I have to adapt the teaching of some teachers and the life to live with somebody.

MT translation: 我必須使一些教師的教導和與某人同住的生活適合 (“I must make the teaching of some teachers and the life with somebody suitable”)

Revised sentence: The teaching of some teachers and the life to live with somebody are waiting there for me to adapt.

MT translation: 一些教師的教導和與某人同住的生活正在那裏等我適應 (“The teaching of some teachers and the life with somebody are waiting there for me to adapt

to”)

Reason: The original translation was not the intended meaning.

In addition, from observing these examples, it can be surmised that MT as a grammar checker can help the learner in the following ways:

- Flagging a wrong word or phrase
- Calling attention to aspects of punctuation
- Drawing attention to polysemy
- Inviting learner to experiment with structures
- Making learner more aware of their interlanguage.

In the follow-up interview, all students except one or two admitted that MT is useful to them in some ways. Most felt it is good for finding translation of individual words or phrases in L2. While about half of the learners thought it was useful to run their English composition through the MT before handing it out, one student suggested that a possible way she would use the MT was to input a Chinese sentence which she found particularly difficult to express in English, and post-edit the English translation to get a satisfactory sentence. While the learners might not have realised it, the pre-editing exercise not only familiarises them with the current MT technology, but also helps them understand more about their competence in English.

## 5. Conclusion

The author has taught both post-editing and pre-editing skills to trainee translators regarding MT, as well as other computer-related translation modules such as translation memory systems. Never

before had the author witnessed the zeal expressed by the students on the same scale as when they were doing the pre-editing task. In the face of the relatively immature MT output, it seems so natural for students to assume the role of a more advanced learner. In doing so, students regain their confidence in the foreign language and many useful self-discovery procedures ensue. What's more, the manipulation of input and the observation on the MT output enable learners to examine the language and the translation skills in a metaphysical way, which is likely to facilitate the generation of new knowledge. In a word, pre-editing is arguably one of the most exciting and effective ways of learning to use MT, learning to translate, and learning the language.

## References

- Cook, Guy. 1998. "Use of translation in language teaching". in Baker, Mona. (ed). *Routledge Encyclopaedia of Translation Studies*. pp. 277-280. London: Routledge.
- Lebauer, Roni. and Robin Scarcella. 1993. *Reactions: Multicultural Reading-Based Writing Modules*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Shei, C-C. 2002. *Combining Translation into the Second Language and Second Language Learning: An Integrated Computational Approach*. Ph.D. Thesis. Division of Informatics, University of Edinburgh.

## Acknowledgement

This is a part of the author's project regarding a computational approach to teaching translation and foreign languages, sponsored by the National Science Council, Taiwan. Project No. 902520S309001. NSC also kindly sponsored my trip to the EAMT Workshop.