

Designing Controlled Chinese Rules for MT Pre-Editing of Product Description Text

Ying Zheng, School of Foreign Languages, Chaohu University, China*

Chang Peng, Anhui Branch National Computer Network Emergency Response Technical Team/Coordination Center of China, China

Yuanyuan Mu, School of Foreign Languages, Chaohu University, China

ABSTRACT

The study aims to investigate how pre-editing based on controlled Chinese rules can be an effective approach to improving Chinese-to-English machine translation (MT) output. Based on the analysis of comparable texts, and by considering the rules of modern business writing and the differences in sentence structure between Chinese and English languages, four controlled Chinese rules for product description text are proposed: 1. Every sentence should have an explicit subject; 2. There are no repetitive complimentary expressions; 3. Sentences should be short; and 4. The Chinese sentence structure should be complete, with clear logical relationships within and between sentences. In accordance with the four rules, five corresponding pre-editing methods are introduced. Then, taking Xiaomi Air2 SE earphones' product description as an experimental text, the study examines the influence of pre-editing in accordance with controlled Chinese rules on MT output quality. The results show that such pre-editing can significantly improve MT output in dimensions of adequacy, fluency, and style.

KEYWORDS

Adequacy, Comparable Texts, Fluency, Logic, Repetition, Sentence Structure, Style, Subject

INTRODUCTION

This paper focuses on the pre-editing methods in accordance with controlled Chinese rules to help machines deliver better output in Chinese-to-English (C-E) translation. Neural machine translation (NMT) has achieved great progress in accuracy and fluency, but in C-E translation, due to the huge differences between the two languages and the cultures behind the languages, MT (specifically referring to NMT in this paper) cannot yet deliver high-quality output for operative texts, for instance, product description text. Product description needs to provide information and attract customers, and it is also updated frequently. So, the translation of such texts should be efficient and high-quality. To achieve the balance between quality and efficiency, pre-editing and post-editing of MT are necessary. This study mainly focuses on pre-editing methods for MT.

According to the standards for post-editing of MT output, "Pre-editing refers to modifying the source language content before machine translation (MT) to facilitate the process, improve raw

DOI: 10.4018/IJTIAL.313919

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

translation output quality, and therefore reduce the post-editing workload, especially if one document is to be machine translated into several languages” (International Organization for Standardization or ISO, 2017). Product descriptions in Chinese are often translated into many languages, and effective pre-editing can greatly reduce the post-editing workload (Feng & Gao, 2017). Based on an experiment, this study aims to address the following questions: Can controlled Chinese be applied to non-technical texts? If yes, for the product description text, what are the pre-editing methods in accordance with controlled Chinese rules? Can such pre-editing methods really improve MT output quality? In what dimensions and to what extent can pre-editing methods improve raw translation output quality?

The next section introduces the previous studies on CL and pre-editing based on CL. Section 3 infers rules of controlled Chinese for product description text and corresponding pre-editing methods from comparative analysis. Then, Section 4 conducts an experiment with such rules and methods and analyzes the pre-editing methods in detail with examples, while Section 5 discusses the results of the experiment. Finally, Section 6 comes to conclusions and discusses future research directions.

PRE-EDITING BASED ON CONTROLLED LANGUAGE

Controlled language (CL) refers to “subset of natural languages whose grammars and dictionaries have been restricted in order to reduce or eliminate both ambiguity and complexity”(ISO, 2017). Controlled language, commonly controlled English, has been mainly applied to technical documents including technical guides and maintenance manuals in technology industries such as aviation since the 1970s, with the aim of improving the readability and translatability of technical texts. Since 2003, the research on CL has turned to controlled translation, i.e., the combination of CL and MT. CL can effectively eliminate the ambiguity of the source language content, which is the biggest challenge for MT, greatly improve the accuracy of MT and reduce post-translation editing workload (Yuan, 2003). Studies on CL have focused on translation between European languages (such as English, German and French), and the results show a generally positive impact of CL rules on the MT output (Bernth & Gdaniec, 2001; Reuther, 2003; Marzouk, 2021). Marzouk’s recent study (2021) indicates that neural MT offers a promising solution that no longer requires CL rules for improving the MT output in the context of technical texts. But we wonder if the idea of CL can be introduced to the MT of non-technical texts, and we aim to answer that question in this paper.

Pre-editing based on CL to make the content suitable for MT is also part of a controlled translation environment (Torrejón & Rico, 2002). Chinese scholars Feng and Gao studied the influence of pre-editing in accordance with controlled English on MT (Google, English to Chinese) with iPhone 5 English user manual as the experimental text, and found that “pre-editing based on controlled English has a positive influence on the enhancement of both the quality of MT regarding adequacy (information content) and fluency of the target text” (Feng & Gao, 2017). Yuan (2003) and Liu (2014) discussed the benefits of CL and pointed out that controlled English can help improve E-C MT of technical texts in terms of quality and efficiency. As for controlled Chinese, Hu put forward the idea of controlled C-E translation and gave ten suggestions, which provide important insights for pre-editing in accordance with controlled Chinese rules (Hu, 2005). Overall, the study on controlled Chinese has not attracted much attention, although controlled Chinese may have considerable potential in non-technical fields.

CONTROLLED CHINESE RULES AND PRE-EDITING METHODS FOR PRODUCT DESCRIPTION TEXT

In China’s cross-border e-commerce, Chinese to English (C-E) MT is employed frequently. The aim of controlled Chinese or pre-editing in accordance with controlled Chinese rules is to help improve C-E MT in terms of adequacy, fluency, and style. Based on the analysis of comparable bilingual texts, and by considering the rules of modern business writing and the differences between Chinese and English languages, we propose controlled Chinese rules and corresponding pre-editing methods for product

description text. The comparable texts selected in this study are: the English text, which includes the Overview and Highlights in the Product Information of AirPods with Charging Case on Apple's official website¹; the Chinese text, which comprises product description (excluding notes) of Mi True Wireless Earphones 2 Basic (named Air 2 SE in China) on product detail page on Xiaomi's official website²; the English translation text, which is the English version of the Chinese text on Xiaomi's UK official website³. Xiaomi's revenue from overseas markets increased by 33.7% to 163.6 billion yuan in 2021, indicating that Xiaomi products are gaining popularity in the international market. The success of Apple products in the international market means that its writing style can be learned from. The earphones of the two brands have similar functions, so the product features introduced are similar. Therefore, the three texts above can be qualified as comparable texts. It should be noted that the selected texts do not contain technical texts such as packing list, parameters, compatibility, or notes on the product detail page, as MT can already deliver satisfactory outputs in the domain of technical texts and may not need pre-editing. After a comparative analysis of the three comparable texts, the writing features of the English text are summarized. Based on such features as well as business writing rules and the differences in sentence structure between Chinese and English languages, the controlled Chinese rules and corresponding pre-editing methods are put forward.

Rule 1—Every Sentence Should Have an Explicit Subject

The English text only uses active voice; the subject is “you” (39.1%), or the product name AirPods (43.5%), or a certain technology (17.4%)⁴. Using you as the subject can give the text a conversational tone; using product name as the subject can reinforce customers' impression of the product; and using a certain technology as the subject can show the product's technological advantage. For instance, the sentence *...and when you're on a call or talking to Siri, an additional speech-detecting accelerometer works with beam forming microwaves to filter out external noise and focus on the sound of your voice.* uses *you* and *an additional speech-detecting accelerometer* (a specific technology) as subjects. The adoption of active voice is also in line with the principle of modern business writing. As Taylor mentioned, “The use of passive voice will create a distance between writers and readers...Today's business writing should adopt the active voice more, so that it will be more positive and clearer” (2012/2014, p. 29). The Chinese text also uses the active voice, but the subject is often omitted, which is a prominent feature of the Chinese language. When processing Chinese sentences with no explicit subject, MT would translate them into the passive voice, or add the impersonal pronoun *it* as the subject, as shown in Example 1.

Example 1

Source Text (ST): 采用142mm大尺寸的复合振膜动圈单元,声音更加丰满浑厚,听感自然,还原更多真实质感和细节。同时支持AAC音频编解码技术,带来听得到的音质提升与体验,尽享动听。

MT (Sogou Translate): The moving coil unit with a large size of 142mm is adopted, which makes the sound fuller and more natural, and restores more real texture and details. At the same time, it supports AAC audio coding and decoding technology, which brings about the improvement and experience of the audible sound quality, and enjoys pleasant sounds.

Official Version: The dynamic driver with a 142mm large composite diaphragm is used to make the sound fuller and richer with a more natural feel, restoring more real texture and details. At the same time, AAC audio codec technology is supported to improve the audio quality and experience, providing users with clear sounding audio.

We can see from Example 1 that to translate the Chinese sentence with no subject, MT adopts the passive voice, or uses *it* as the subject, while the official English translation version directly applies the passive voice. The use of passive voice is not a translation error, but it is too formal and does not

suit the business context. Since MT can't identify the subject of the source language content, using *it* as the subject replacement might cause ambiguity. Therefore, in Chinese writing or pre-editing, if we can make sure each source sentence has an explicit subject, the translation accuracy will be improved, and passive voice frequency reduced. In Example 1, we can add the product name as the subject. Therefore, controlled Chinese Rule 1 can be determined: Every sentence should have an explicit subject. The corresponding pre-editing method is: Add a subject.

Rule 2—There Should Be No Repetitive Complimentary Expressions

The English text has 304 words, the Chinese text 717 Chinese characters, and the English translation text 536 words. Customers can use less time to browse the English text. One of the reasons for the difference in text length is the repetitive complimentary expressions in the Chinese text. In the English text, complimentary expressions with similar meanings are not repeated. In contrast, the Chinese text employs a variety of complimentary expressions with similar meanings to enhance the effect and highlight the advantages of the product, which is a typical feature of the product description by Chinese brands. For instance, the following Chinese sentence introduces the sound quality, 采用142mm大尺寸的复合振膜动圈单元,声音更加丰满浑厚,听感自然,还原更多真实质感和细节 (literal translation: Adopting the 142mm large-size composite diaphragm, the sound is fuller and richer, the hearing sense is natural, and restores more real texture and details). The repetitive complimentary descriptions of sound quality do not contain more information and will not in fact highlight the product advantage. In contrast, similar information in the English text is contained in just one phrase *Rich, high-quality audio and voice* in the Highlights section, which is more concise and even contains more information (emphasizing the quality of *voice* in calls). Hence this type of expressions can be reduced in the pre-editing of the Chinese text. Based on the preference for concise business language, controlled Chinese **Rule 2** can be established: There should be no repetitive complimentary expressions. The corresponding pre-editing method is: Remove repetitive complimentary expressions.

Rule 3—Sentences Should Be Short

The average sentence length of the English text is 12.16 words per sentence, that of the Chinese text is 27.58 Chinese characters per sentence, and that of the English translation text is 13.74 words per sentence (see the basic statistics of the three texts in Table 1). The sentence count of the Chinese text (26) is similar to that of the English text (25), but the average sentence length (25.78) is significantly higher than that of the English text (12.16). Even when we take into account the ratio of 1:1.73 which exists in English-Chinese parallel corpora (Li, 2021), the difference is still significant. The English translation divides the Chinese sentences, increasing the sentence count (to 39) and decreasing the sentence length (to 13.17), which makes the translation version relatively easier to understand. However, the Flesch Reading Ease of the translation version is lower than that of the English text, and its Flesch-Kincaid Grade Level⁵ higher, which means the translation version still requires more effort to understand. The main factors affecting the reading difficulty of Flesch Reading Ease and Flesch-Kincaid Grade Level are word length and sentence length, which indicates that we should choose short words and cut sentences to reduce the reading difficulty when translating such texts into English. In the product description by Chinese brands, sentences can be long, each consisting of several comma-connected short clauses. The short clauses connected by commas are easy for Chinese readers to understand. However, if such long sentences are machine-translated into English, the MT output will be long sentences, sometimes with confusing structures, which can slow down the reading speed and then lose readers (see Table 2). Taylor pointed out that “short sentences are best” in business writing (2012/2014, p. 47). In order to adapt to the reading habits of the target customers and achieve the purpose of effectively transmitting information, controlled Chinese **Rule 3** can be determined: Sentences should be short. The corresponding pre-editing method is: Cut sentences short.

Table 1. Basic statistics of the three comparable texts

Text type	Word count	Sentence count	Words per sentence(avg)	Syllables per word (approx.)	Flesch Reading Ease	Flesch-Kincaid Grade Level
The English text	304	25	12.16	1.63	56.74	8.37
The Chinese text	717	26	27.58	/	/	/
The English translation text	536	39	13.74	1.72	47.04	10.11

Table 2. The influence of sentence word count on its reading ease (Taylor, 2012/2014, p. 47)

Word count of a sentence	The proportion of people who can understand the meaning of the sentence the first time they read it.
7 to 10 words	95%
5 to 20 words	75%
27 or more words	4%

Rule 4—The Chinese Sentence Structure Should Be Complete, With Clear Logical Relationships Within and Between Sentences

Between Chinese and English languages, there is one prominent difference in sentence structure: the Chinese language often omits the content that can be understood according to the context, so the sentence structure may not be complete or may not include conjunctions or adverbs which indicate the relationships within and between sentences. In other words, the logic within and between sentences may not be expressed in words or phrases. In contrast, although the English language has omissions on certain occasions, its sentence structures are complete and the logic within and between sentences is expressed explicitly in words or sentence structure. To enable the MT to deliver complete English sentences with clear logic in C-E translation, **Rule 4** of controlled Chinese can be determined: The Chinese sentence structure should be complete, with clear logical relationships within and between sentences. The corresponding pre-editing method is: Add omitted words and make explicit logical relationships within and between sentences. This method can help MT better grasp the semantic content and logic of ST and deliver better outputs.

Table 3 is a summary of the controlled Chinese rules and the corresponding pre-editing methods for the product description text.

Table 3. Controlled Chinese Rules and Corresponding Pre-editing Methods

Controlled Chinese Rules	Pre-editing Methods
1 Every sentence should have an explicit subject.	Add a subject.
2 There should be no repetitive complimentary expressions.	Remove repetitive complimentary expressions.
3 Sentences should be short.	Cut sentences short.
4 The Chinese sentence structure should be complete, with clear logical relationships within and between sentences.	Add omitted words.
	Make explicit logical relationships within and between sentences.

CASE STUDY

Experiment Design

The translation engine selected in this study is Sogou Translate based in China. Sogou Translate is one of the mainstream generic C-E translation engines and has received good comments from users. According to Cai Xinjie and Wen Bing's case study on MT of publicity texts, the comprehensive performance of Sogou Translate in C-E translation is better than that of Youdao Translate, Baidu Translate and Google Translate (2021).

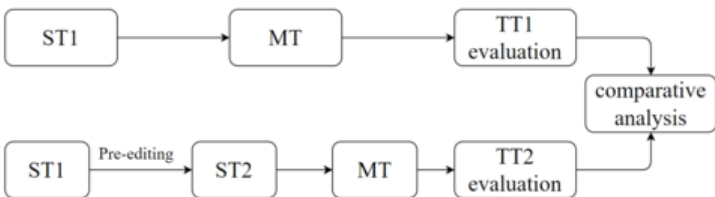
Firstly, we entered the Chinese source text (ST1) directly in Sogou Translate to get target text 1 (TT1) as the control group. Secondly, ST1 was pre-edited in accordance with the above-mentioned controlled Chinese rules; the pre-edited text was called ST2; then we entered ST2 in Sogou Translate to get TT2 as the experimental group; To imitate real-life translation practice, we gave TT2 light post-editing, which means using the raw MT output as much as possible to ensure efficiency, and editing any inappropriate content (ISO, 2017). Thirdly, the translation qualities of the two groups were evaluated against the same criteria. Ten professional translators were invited to grade control group and experimental group on a 10-point Likert Scale. Finally, the scores of the two groups were compared to see whether the translation quality of the experimental group was improved. The scores were uploaded to the online data analysis platform SPSSAU to analyze the influence of pre-editing in accordance with controlled Chinese rules on MT output quality. The procedure is shown in Figure 1.

Since pre-editing in accordance with controlled Chinese rules may remove unnecessary content (such as repetitive complimentary expressions), thus reducing on the superficial level the accuracy of MT, the frequently used BLEU method was not adopted for evaluation. The flexible and accurate human evaluation was applied. In the aspect of evaluation criteria, since there are no recognized evaluation criteria for pre-editing, and the purpose of both pre-editing and post-editing is to improve the quality of MT, we decided to use the criteria based on the requirements for post-editing in ISO's standard (2017): "The output should be accurate, comprehensible and stylistically adequate". Hence the criteria in this study include three dimensions: adequacy, fluency, and style: 1. Does the translation contain adequate information that needs to be conveyed? 2. Is the translation fluent and easy to understand? 3. Does the translation observe the writing style of the modern business world? The criteria are similar to those adopted by Professor Li Mei in her research on the impact of source texts on human translators in the process of MT post-editing (Li Mei 2021).

Examples of Pre-Editing Methods

In the following part, we will give examples in which pre-editing methods in accordance with controlled Chinese rules are applied. Each example will focus on one pre-editing method, although in practice several methods might be applied at the time to the same sentence. The purpose of pre-editing is to improve the quality and efficiency of MT, so in the pre-editing step, we made only the necessary changes to the ST1 to ensure efficiency. If necessary, light post-editing was applied to TT2 on the terminology level, but full post-editing on the style level was not applied.

Figure 1. Experiment Procedure



Pre-Editing Method 1

Add a subject. Chinese grammar permits an independent clause to lack an explicit subject. Readers can understand the clause with a null subject according to the context. But the machine, for the time being, can't always accurately identify the implied subject based on the context. When the null subject phenomenon appears in the ST1 in Example 2, the machine uses *it* as the subject (in TT1) or adopts the passive voice (in the official version). *It* is not an appropriate subject since we should use the plural form for earphones, and the passive voice is not suitable for business writing. In the pre-editing process, we added the subject 耳机(*earphones*) and 你(*you*). We also changed the second comma to a full stop to cut the long sentence into two short ones. TT2 shows that after pre-editing, the MT output has a clear subject and adopts the active voice. Overall, the adequacy and fluency of the MT output are improved.

Example 2

ST1: 采用高灵敏电容式入耳佩戴检测,可检测佩戴状态,音乐播放过程中摘下耳机即可暂停音乐,重新佩戴恢复播放。

TT1 (original Sogou translation): Using high sensitive capacitive in-ear wear detection, it can detect the wearing state. During the music playing process, you can pause the music by taking off the earphone, and resume playing after wearing it again.

ST2 (after pre-editing): 耳机采用高灵敏电容式入耳佩戴检测,可检测佩戴状态。音乐播放过程中你摘下耳机即可暂停音乐,重新佩戴恢复播放。

TT2 (Sogou translation after pre-editing): Headphones adopt highly sensitive capacitive in-ear wear detection, which can detect the wearing state. When playing music, you can pause the music by taking off your headphones, and resume playing by wearing them again.

TT3 (after post-editing): The earphones adopt highly sensitive capacitive in-ear wear detection, which can detect the wearing state. When playing music, you can pause the music by taking off your earphones, and resume playing by wearing them again.

Official version: Highly sensitive capacitive in-ear wearing detection is used to detect wearing status. While listening to music, pause by taking out the earphones and resume by putting them back in.

To keep the terms consistent, *headphones* should be changed to *the earphones* to refer to the product in question in the light post-editing step. However, we suggest using the product name as the subject (e.g., AirPods as the subject in the English text). But the product's English name *Mi True Wireless Earphones 2 Basic* is too long to be used as a frequent subject. If the Mi product has a short English name to replace the generic term “the earphones”, the repetition of the product name as the subject can deepen customers' impression of the product and enhance the effect of product publicity.

Pre-Editing Method 2

Remove repetitive complimentary expressions. In Chinese language, complimentary expressions with similar meanings often appear together to enrich the expressions and boost the effect. According to controlled Chinese Rule 2, we should remove the repetitive content to make the message more efficient. In Example 3, 带来听得到的音质提升与体验,尽享动听 (*Bring audible sound quality improvement and experience, and enjoy the sound beauty*): the Chinese sentence structure is complex; 听得到的(*audible*) should not modify 体验(*experience*), and 尽享动听(*enjoy the sound beauty*) is a repetitive expression. The clause can be shortened to 带来更好的音质与体验 (*bring better sound quality and experience*), and the sentence subject 耳机(*earphones*) can be added. With pre-editing, ST2 is more concise and coherent, and so is the TT2. In light post-editing, to keep terms consistent, we replaced *headphones* with *the earphones* and *sound* with *audio*. In Example 4, the second clause 一触随心而控 (*one touch and control at will*) is also an unnecessary compliment, and thus can be removed; we post-edited the term *touch manipulation* to *tap control*. Comparing the TT3s with the

official versions in Example 3 and Example 4, we can see that MT integrated with human effort can sometimes deliver better output than the official version.

Example 3

ST1: 同时支持AAC音频编解码技术,带来听得到的音质提升与体验,尽享动听。

TT1: At the same time, it supports AAC audio coding and decoding technology, which brings about the improvement and experience of the audible sound quality, and enjoys pleasant sounds.

ST2: 同时耳机支持AAC音频编解码技术,带来更好的音质与体验。

TT2: At the same time, headphones support AAC audio codec technology, which brings better sound quality and experience.

TT3: At the same time, the earphones support AAC audio codec technology, which brings better audio quality and experience.

Official version: At the same time, AAC audio codec technology is supported to improve the audio quality and experience, providing users with clear sounding audio.

Example 4

ST1: 触摸操控,一触随心而控

TT1: Touch control, one touch at will

ST2: 触摸操控

TT2: Touch manipulation

TT3: Tap control

Official version: Tap to operate, and control at your own leisure with a single touch

Pre-Editing Method 3

Cut sentences short. The Chinese language tends to connect clauses with a comma, and often omits the subject as mentioned above, making it difficult for MT to recognize the sentence structure and find the right subject for each Chinese clause. We can cut sentences short with a semicolon (for closely-related clauses) or a full stop (for loosely-related clauses) to help MT better understand the sentence structure. ST1 in Example 5 is a shortened sentence from the original Chinese text to illustrate the importance of cutting sentences. In ST1, the subject of the first clause 真无线设计,打破主从限制 (*True wireless design, breaking the master-slave restriction*) is different from that of the second clause 左右耳机、单双耳机灵活切换 (*Flexible switching between left and right earphones, single and double earphones*): the subject of the first clause is 真无线设计 (*True wireless design*), and there is no explicit subject in the second clause. Since the Chinese clauses are joined by a comma, the machine chose 真无线设计 (*True wireless design*) as the subject for the whole sentence and translated it into ...*design, breaking..., switching....* The machine used the subject of one clause as the subject for other clauses with a null subject, which is not accurate in translation. Such mistakes are quite common in C-E MT. To avoid such mistakes, we can cut sentences short with a full stop or semicolon, so that the machine can better analyze the sentence structure. As for example 5, the two clauses in ST1 are logically close-related, so we can use a semicolon to cut the sentence. On certain occasions, we need to use a full stop to separate loosely-related clauses. TT2, with the use of a semicolon, is more accurate in logic although in this case the change of punctuation mark hasn't brought about significant improvement. In addition to cutting sentences, we can also add certain words to indicate the implicit logic within and between Chinese clauses, which can help the machine deliver more coherent output. This method of making explicit sentence relationships will be talked in detail in 4.2.5.

Example 5

ST1: 真无线设计,打破主从限制,左右耳机、单双耳机灵活切换。

TT1: True wireless design, breaking the master-slave restriction, flexible switching between left and right earphones and single and double earphones.

ST2: 真无线设计,打破主从限制;左右耳机、单双耳机灵活切换。

TT2: True wireless design, breaking the master-slave restriction; Flexible switching between left and right earphones and single and double earphones.

Pre-Editing Method 4

Add omitted words. The Chinese language also often omits prepositions, conjunctions or other function words that can show the sentence logic, and even omits certain content words when the context can help the reader easily get the message. The conciseness of the Chinese language might bring problems to C-E MT—the machine cannot accurately infer the omitted content. So, in the pre-editing step, we can add such omitted words to help the machine recognize the logic as well as the semantic content of the ST and deliver accurate output. TT1s in Example 6 and Example 7 indicate that the machine didn't recognize the logic of the ST and gave a word-for-word translation. In ST2 of Example 6, the preposition 在...中(*during*) and the subject 你(*you*) were added; in ST2 of Example 7, the subject 你(*you*), the preposition ...时(*when*), and the verb 出现(*appear*) were added. Such additions were meant to make the semantic content of ST complete, and the subject and logic clear. We can see from the TT2s in both examples that, after pre-editing, the performance of MT is better in terms of adequacy and fluency. Still, we admit that the TT2s are not perfect and need to be improved in writing style for publicity, a task to be done in full post-editing. The official version, adopting the third person and passive voice, can also be refined in style.

About the sequence of subject, predicate and object in Chinese, that is, 出现弹窗(*appears pop-up window*) or 弹窗出现(*pop-up window appears*), 智能连接你看得到(*smart connection you can see*) or 你看得到智能连接(*you can see smart connection*), an experiment was carried out by entering different sequences into the translation engine. The results show that there is no difference in MT output: the subject, predicate and object were all recognized correctly and reorganized according to the English norm. It means that when the relationship between Chinese subject, predicate and object is clear, the machine can correctly recognize the semantic content and express the meaning according to the English norm. Hence in such situations, the word order of the ST does not need to be changed, thus reducing the workload of pre-editing.

Example 6

ST1: 双麦克风通话降噪 无惧干扰

TT1: Dual-microphone call noise reduction without fear of interference.

ST2: 双麦克风在通话中降噪,你无惧干扰

TT2: Dual microphones can reduce noise during the call, so you are not afraid of interference.

Official version: With dual microphones, noise during calls is reduced, making sure there is no disturbance.

Example 7

ST1: 开盒弹窗,智能连接看得到

TT1: Open the box pop-up window, which can be seen by smart connection

ST2: 你开盒时出现弹窗,智能连接你看得到

TT2: Pop-up window appears when you open the box, and you can see the intelligent connection.

TT3: Pop-up window appears when you open the case, and you can see the smart connection.

Official version: A window pops up when the case is opened, and the smart connection becomes visible.

Pre-Editing Method 5

Make explicit logical relationships within and between sentences. The Chinese sentence may not include conjunctions or adverbs which indicate the relationships within and between sentences. Readers can infer the logic from the context, but the machine cannot achieve that at present. So, we can make explicit the logical relationships in ST to help the machine understand ST better and produce more logical TT. In Example 8, the first clause 采用全新架构对软硬件进行深度优化 (*using a new architecture to deeply optimize the software and hardware*) is the cause of the second clause 左右耳同步连接 (*the left and right ears are synchronously connected*), but they appear as coordinate clauses joined by a comma. So TT1 also presents coordinate clauses joined by *and*, which can't reflect the causation relationship between two clauses. In the pre-editing step, we added 实现 (*realize/achieve*) before the second clause to make the causation relationship explicit. We also added the subject 我们 (*we*) and replaced 耳 (*ear*) with 耳机 (*earphones*) to avoid ambiguity. With the input of ST2, the MT output TT2 has a clearer logic and adopts the preferred active voice. In Example 5, we can also use this method. If we replace the semicolon with 实现 (*realize/achieve*), the whole sentence will be machine-translated into *True wireless design breaks the master-slave restriction and realizes flexible switching between left and right earphones and single and double earphones*, suggesting the machine can better recognize the logic of ST after the pre-editing step.

Example 8

ST1: 采用全新架构对软硬件进行深度优化,左右耳同步连接。

TT1: The software and hardware are deeply optimized by adopting a brand-new architecture, and the left and right ears are connected synchronously.

ST2: 我们采用全新架构对软硬件进行深度优化,实现左右耳机同步连接。

TT2: We use a new architecture to deeply optimize the software and hardware and realize the synchronous connection between left and right headphones.

The translation engine used in this study is for general purposes and has not been trained with data for the specific domain, so translation errors in terminology frequently appear. Presently, there are two solutions to the problem: custom MT (MT trained with specific data); or applying terminology database in the post-editing step. Chinese scholars put forward the pre-editing method of “translating terms in advance” for English-Chinese translation (Feng & Gao, 2017), which can ensure the correctness and consistence of terminology. However, for C-E MT, it is discovered in the experiment that, the translation of terms in advance often leads to the machine being unable to deliver correct grammatical forms in English (including tenses and singular and plural word forms), thus increasing the workload of post-editing. Therefore, custom MT or the use of terminology database in post-editing may better solve the terminology problem for C-E MT in specific fields.

Experiment Results

From the above case analysis, we can see that pre-editing in accordance with controlled Chinese rules can improve MT output in the aspects of adequacy, fluency, and style. In order to further examine the effectiveness of controlled Chinese rules, we randomly chose another nine sentences for the experiment. Ten professional translators were invited to evaluate the translation quality against the

same criteria on a 10-point Likert Scale without being told the difference between TT1 (MT output without pre-editing) and TT2 (MT output with pre-editing). The average scores are shown in Table 4.

Paired t-tests were performed to compare the scores of TT1 and TT2 on the SPSSAU online data analysis platform, and the results were as follows: There was a significant difference in adequacy between TT1(M=6.44, SD=0.73) and TT2(M=8.83, SD=0.46); $t(8)=-9.288$, $p=.000$. There was a significant difference in fluency between TT1(M=6.37, SD=0.82) and TT2(M=8.74, SD=0.42); $t(8)=-10.662$, $p=.000$. There was a significant difference in style between TT1(M=6.24, SD=0.63) and TT2(M=8.58, SD=0.50); $t(8)=-10.359$, $p=.000$. We also carried out paired t-tests to compare the value differences in adequacy, fluency, and style, and found that there was no significant difference ($p > 0.05$ in all cases), which indicates that pre-editing can improve adequacy, fluency, and style, but there is no significant difference between the three dimensions. Although the average scores for TT2 are significantly higher than those for TT1, TT2 scores are all in the range of 8.5-8.9, which means that post-editing is still necessary to achieve high-quality translation.

It should be noted that, although the average scores for TT2 are significantly higher than those for TT1, in two questionnaires, the scores for TT2 are equal to, or even slightly lower than those for TT1 in fluency and style dimensions. The evaluators who gave such scores were interviewed about the reasons for their choice. They explained that advertising discourse does not need complete sentences, and that noun phrases are more concise and efficient in conveying information. Therefore, we further studied the product description texts of similar earphones on Amazon.com and have concluded some advice for post-editing: while post-editing product information text, the subject can be removed on the premise that the context is clear and the removal will not cause any misunderstanding; attention should be paid to the parallel structure for a list of product features.

CONCLUSION

Based on the analysis of comparable texts, and by considering rules of modern business writing and the differences in sentence structure between Chinese and English languages, this paper puts forward four controlled Chinese rules and five corresponding pre-editing methods for product description text and discusses the influence of pre-editing in accordance with controlled Chinese rules on MT output quality. The study shows that pre-editing in accordance with controlled Chinese rules can significantly improve MT output quality in dimensions of adequacy, fluency, and style. However, in

Table 4. Human evaluation of MT output with or without pre-editing

	TT1 (without pre-editing) average score			TT2 (with pre-editing) average score		
	Adequacy 1	Fluency 1	Style 1	Adequacy 2	Fluency 2	Style 2
1	5.83	5.67	5.83	8.67	8.17	7.67
2	5.67	5.33	5.33	9	8.83	8.5
3	6	6.83	6.5	8.83	8.5	8.17
4	6	5.5	5.83	9	8.5	8.83
5	6.83	6.67	6.33	7.83	8.33	8.83
6	6	5.67	6.17	8.5	8.67	8.17
7	6.67	7.17	7.33	9.17	9.33	9.17
8	7.83	7.5	5.83	9.17	9.33	9.17
9	7.17	7	7	9.33	9	8.67
Mean	6.44	6.37	6.24	8.83	8.74	8.58

order to produce a translated text of high quality, post-editing is still needed to further improve MT output. It should be noted that the need for pre-editing is related to the language quality of the ST and the style differences between ST and TT. The Chinese writing style for product description discussed in this paper is quite different from the English writing style for the same genre, so controlled Chinese can significantly improve MT output. If the ST has high language quality and is in a similar style to that of the TT, MT can deliver high-quality output without pre-editing.

For controlled English, CL checkers can check whether a text conforms to the writing rules of controlled English. In the future study, we can develop controlled Chinese checkers to help translators check whether the text conforms to the controlled Chinese rules. This paper focuses on controlled Chinese for product description. Future research can be carried out in controlled Chinese for other non-technical texts. Controlled Chinese for general purposes and for specific purposes can be put forward to help improve C-E MT output in various domains. We are aware that NMT based on deep learning is becoming smarter, and the way we interact with machine needs to be continuously updated in accordance with the constant enhancement of MT engines. Only by doing so can we achieve high-efficiency and high-quality human-machine interaction.

COMPETING INTERESTS

All authors of this article declare there are no competing interest.

FUNDING AGENCY

This research was supported by Hefei Social Science Fund [Project No. HFSKQN202238: Research on the Application of Controlled Chinese in Business Text Translation from the Perspective of Human-Machine Interaction]; and Anhui Provincial Department of Education [Project No. 2020wyxm141: Digital Humanities Course Development for Foreign Language Majors from the Perspective of New Humanities] 基金项目: 合肥市哲学社会科学规划项目(课题编号:HFSKQN202238,人机交互理念下受控汉语在商务文本英译中的应用研究);安徽省省级质量工程项目(课题编号:2020wyxm141,新文科视野下的外语专业数字人文教育课程建设).

REFERENCES

- Bernth, A., & Gdaniec, C. (2001). MTranslatability. *Machine Translation*, 16(3), 175–218. doi:10.1023/A:1019867030786
- Cai, X., & Wen, B. (2021). 汉译英机器翻译错误类型统计分析——以外宣文本汉译英为例 [Statistical analysis on types of C-E machine translation errors: A case study on C-E translation of publicity text.]. *Journal of Zhejiang Sci-Tech University*, 46(2), 162–169.
- Feng, Q., & Gao, L. (2017). 基于受控语言的译前编辑对机器翻译的影响 [On the influence of pre-editing in accordance with controlled language on machine translation.]. *Contemporary Foreign Languages Studies*, (2), 63–68, 87, 110.
- Hu, Q. (2005). 机器翻译中的受控语言 [Controlled language in machine translation.]. *Chinese Science & Technology Translators Journal*, 18(3), 24–27.
- International Organization for Standardization. (2017). *Translation services—Post-editing of machine translation output—Requirements* (ISO Standard No. 18587:2017). <https://www.iso.org/standard/62970.html>
- Li, M. (2021). 机器翻译译后编辑过程中原文对译员影响研究 [On the impact of source text on human translators in MTPE.]. *Foreign Language Education*, 42(4), 93–99.
- Liu, Y. (2014). 控制性语言对提高技术规范文本机器翻译质量的实证研究 [Empirical study of controlled language for improving the quality of the machine translation technical specifications.]. *Journal of Chongqing University of Technology (Social Science)*, 28(3), 96–101.
- Marzouk, S. (2021). An in-depth analysis of the individual impact of controlled language rules on machine translation output: A mixed-methods approach. *Machine Translation*, 35(2), 167–203. doi:10.1007/s10590-021-09266-0
- Reuther, U. (2003). Two in one—can it work? Readability and translatability by means of controlled language. In EAMT Workshop: *Improving MT through other language technology tools: resources and tools for building MT*.
- Taylor, S. (2014). *Model Business Letters, E-mails & Other Business Documents*, (7th ed.). (Y. Lu & R. Bai, Trans.). Foreign Language Teaching and Research Press.
- Torrejón, E., & Rico, C. (2002). Controlled translation: a new teaching scenario tailor-made for the translation industry. In *Proceedings of the 6th EAMT Workshop: Teaching Machine Translation*.
- Yuan, Y. (2003). 受控语言:机器翻译的新途径 [Controlled language and machine translation.]. *Shanghai Journal of Translators for Science and Technology*, (3), 77–80.

ENDNOTES

- ¹ Website: <https://www.apple.com/shop/product/MV7N2AM/A/airpods-with-charging-case>; Access date: July 29, 2021.
- ² Website: https://www.mi.com/buy/detail?product_id=12036; Access date: July 29, 2021.
- ³ Website: <https://www.mi.com/uk/mi-true-wireless-earphone-2-basic/>; Access date: July 29, 2021.
- ⁴ The data does not include phrases. Only the subjects of clauses (including main clauses, coordinate clauses and subordinate clauses) were calculated. The total count of clauses is 23. Nine clauses have *you* as the subject (in 4 clauses, *you* as the subject is omitted); Ten clauses have product name *AirPods* or *they* (referring to *AirPods*) as the subject; Four clauses use a certain technology as the subject.
- ⁵ Flesch Reading Ease and Flesch-Kincaid Grade Level are widely used readability tests in the United States. For Flesch Reading Ease, the higher the score (1–100), the easier a piece of text is to read. Flesch-Kincaid Grade Level corresponds to the US grade level of education. The lower the grade level, the easier a piece of text is to read. Text intended for the general public should aim for a grade level of around 8, so the text should be understood by 13 to 15-year-olds.