

ChatGPT can Make Unreadable Chinese Texts Even Less Readable

Alex B.-c. Poon

The work described in this paper was fully/substantially/partially supported by InnoHK initiative, The Government of the HKSAR, and Laboratory for AI-Powered Financial Technologies.

September 29, 2023

Contents

1	Motivation	2
2	Methodology and Results	2
3	Inability to Correct Unneccesary Calques	2
3.1	Example - Unneccesary <i>Semantic Calques</i> (‘diplomatic’)	3
3.2	Example - Unneccesary <i>Phraseological Calques</i> (‘the rest is history’)	4
4	Inability to Correct Redundancy	4
4.1	Example - Threefold Redundancy	4
4.2	Example - Paragraph with Both Redundancy and an Unneccesary Calque	5
5	Inability to Correct Confusing Format Switches	5
5.1	Example - Use of Arabic Numerals without Achieving Simplification	5
6	‘False Positives’ which Lead to Semantic Losses	6
6.1	Example - Change of Objects	6
6.2	Example - Meaning Loss & Meaning Shift	6
7	Suggested Solutions	6
8	Conclusion	7

1 Motivation

The readability of chatbot responses not only influences the efficiency of message delivery but also shapes customers' perception. For instance, an unfavourable response from a banking chatbot could be:

「本銀行的服務有能力深度優化閣下公司陳舊的企業機制，除此之外，本行再提供其他提供家庭友善方案。[sic]」

This response is far from ideal due to its excessive length, inclusion of calques, semantic redundancies, and reliance on stock locutions. These readability issues will be further discussed in the upcoming sections. **It is intriguing to see whether ChatGPT can pinpoint Chinese readability issues and correct them, as this will save product designers from reinventing the wheel – to devise a Chinese readability evaluator and sentence generator.**

2 Methodology and Results

We collected 405 short paragraphs in written Chinese with low readability from an archived educative column of the Hong Kong tabloid *am730*, using its online version. Its author, Koo, is a renowned Chinese language expert who had a prolific career as an educator, a university researcher, a writer, and an editor for numerous newspapers and magazines in Hong Kong [1]. The dataset has problematic clauses labelled manually using *Doccano*, and includes suggested corrections.

We created a systematic approach to classify the Chinese readability issues found in the dataset:

- a) Unnecesary calques
 - i) Semantic calques
 - ii) Phraseological calques
- b) Redundancy
- c) Confusion of Chinese terms with close but different meanings
- d) Use of *expletives*, words or phrases used to fill out a sentence or a line of verse without adding to the sense

- e) Format-related issues, e.g., unnecessary use of Arabic numerals

The experiment, conducted in September 2023, was to make ChatGPT (GPT-4 version) point out the clauses with low readability in these paragraphs using the prompt 「請找出以下句子較難理解的地方。」. We not only discovered that ChatGPT was ineffective at identifying readability issues in written Chinese, but also observed that most paragraphs became even less readable.

An overwhelming 89% of the 'corrected' paragraph failed to become more readable, most of which become even less readable. It was found that for 78% of the paragraphs, ChatGPT even lengthened them. The average paragraph length was increased by 134%. The additions were simply semantic overlaps and are not explanations of technical terms. For generations, Chinese has been prized for its conciseness in comparison to other languages. Consequently, paragraph length stands as one of the more significant metrics for assessing Chinese readability.

Through manual inspection, it was also found that roughly 33% of the paragraphs have a loss of meanings by altering keywords. **These findings give rise to concerns regarding the chatbot's Chinese comprehension ability.**

For validation, we repeated the process using other prompts like 「請找出以下句子可讀性低的地方。」, which yielded similar results.

We recognise that randomness and prior conversations can impact response outcomes. Yet, when refreshing the response or initiating a new response from scratch, the results remain similar. Most of the readability issues persist, and in some instances, the 'corrected' paragraph may become even more challenging to comprehend.

3 Inability to Correct Unnecesary Calques

A *calque* is not a usual translation. It refers to 'a word or clause borrowed from another language by literal word-for-word or root-for-root translation' [2].

Most calques are unnecessary. As Koo remarked, **there are many more elegant and understandable ways to express the same concept in Chinese, with numerous usages found centuries ago** [1].

Calque clauses like 「外交」 (as an adjective, from ‘diplomatic’), 「自我成真預言」 (from English ‘self-fulfilling prophecies’), 「用戶友善」 (from English ‘user-friendly’) engender a readability issue, because **if the reader does not know their English etymological sources beforehand, it will be nearly impossible for him or her to understand what they mean**. The older generation, mostly monolingual, is especially unfamiliar with these neologisms.

3.1 Example - Unnecessary Semantic Calques (‘diplomatic’)

There are many examples, so we will just pick a few to illustrate in this paper. We will begin with *semantic calques*, where additional meanings of the source word are transferred to the word with the same primary meaning in the target language.

「李克強的回答十分外交：『不同國家擁有不同歷史文化背景，對人權認識不同。』」 (*Li Keqiang’s answer was very diplomatic, “As different countries have dissimilar historical and cultural backgrounds, their understandings of human rights are different.”*)

ChatGPT edit: 「李克強的回答充滿外交手段，他表示：『不同國家擁有各自獨特的歷史文化背景，因此對人權的理解也各不相同。』」

The phrase suggested by ChatGPT is much longer, and it failed to identify the sentiment of words. It is common for an English word to have different meanings, which is known as a *homonym*. ‘Diplomatic’ may mean ‘pertaining to diplomacy’, but can also be a synonym of ‘smooth’. The calque (better expressions include 圓滑 and 得體) takes the latter sense, which is always positive. According to the Cambridge English Dictionary at least, the term is marked **‘approving’**, defined as ‘acting in a way that does not cause offence’, e.g. “Ask him nicely - be diplomatic.[3]” The Longman Dictionary also corroborates that the term is positive in this sense, defining it as ‘dealing with people politely and skilfully without upsetting them’ [4].

手段 in the suggested edit is actually mildly negative, unlike the positive 手腕 and neutral 手法、策略 etc. It can combine with other negative terms to form negative expressions, such as 不擇手段 (‘by hook or by crook’), 卑劣手段 (‘nasty tricks’) and 高壓手段 (‘strong arm’, ‘iron fist’) [5]. Thus, the meaning of the sentence has totally changed.

For an in-depth diagnosis, we asked ChatGPT more questions. Apart from failing to point out that the clause has low readability, ChatGPT also failed to identify that:

1. The term is positive as per multiple authoritative dictionaries, classifying it as neutral instead:

Question: What is the sentiment of 「外交」 in the sentence 「李克強的回答十分外交：『不同國家擁有不同歷史文化背景，對人權認識不同。』」?

ChatGPT’s answer: 在給定的句子中，描述李克強的回答為「十分外交」，是在強調他的回答具有國際外交上的特徵，即通常采用外交手段，而不是強調其褒義或貶義。[sic] (... *It is neither positive nor negative.*)

2. The term is used as an adjective in this sentence, suggesting it to be a noun instead:

Question: What is the part of speech of 「外交」 in the sentence 「李克強的回答十分外交：『不同國家擁有不同歷史文化背景，對人權認識不同。』」?

ChatGPT’s answer: 「外交」這個詞的詞性是名詞。在給定的句子中，它用作名詞，用來形容或描述李克強的回答方式，指的是一種特定的國際事務和政策協商方法。[sic] (... *The part of speech is ‘noun’.*)

3. The term is a calque of English ‘diplomatic’ in this sentence, suggesting a Chinese word origin instead:

Question: What is the etymology of 「外交」 in the sentence 「李克強的回答十分外交：『不同國家擁有不同歷史文化背景，對人權認識不同。』」?

ChatGPT’s answer: 「外交」這個詞來自中文，其字源可以追溯到古代漢語。這個詞的形成

是由「外」和「交」兩個字組合而成。「外」表示外部、外面的意思，用來表示與國內事務不同的國際事務。「交」表示交往、交流、交涉的意思，用來表示在國際層面上的外部交往和協商。[sic] (... *The etymology is ancient Chinese (Han).*)

3.2 Example - Unnecessary *Phraseological Calques* ('the rest is history')

There is another type of calque called the *phraseological calque* which is derived from idioms of the source language [6], **thus hindering readability even more, since it cannot be interpreted literally in both source and target languages.**

The undesirable expressions 「餘下的是歷史」，「剩下的，就是歷史」 etc. are calques of the English idiom 'and the rest is history' and do not actually pertain to history. Instead, they mean 'everyone knows the rest of a story you have been telling'.

「沒有賈伯斯，就沒有iPod、iPad、iPhone甚至後來的iPhone5，剩下的，就是歷史。」

ChatGPT edit: 「如果沒有賈伯斯，就不會有iPod、iPad和iPhone。後來的iPhone 5也是由他的遺產所啟發的。總之，他的影響力深遠，留下了不可磨滅的歷史。」 (... *Anyhow, he had a longlasting impact, and left an indelible history [sic].*)

ChatGPT's edit has many more issues than the original text. The new paragraph is almost twice as long, which is not a readability improvement. Besides, we would not say 'left a history' (歷史) in both English and Chinese, but 'left a legacy' (偉績). It also introduces redundancies, as 'had a longlasting impact' is identical to 'left an indelible legacy'.

Again, '[Jobs] left an indelible legacy' is not what the author wanted to express. The calque is supposed to mean 'everybody knows the rest of the story (everybody knows what happened after Job's death, suggesting a decline in Apple's influence)'. A suggested edit should be 「後來的事，盡人知之，不在話下」 [1].

Paragraphs that incorporate the same calque produced similar results on ChatGPT, for instance,

「約二零零零年，北京新興的混圈子在舞場音樂退散後，要找個清靜地方歇歇，就自行跑到後海。也差不多在這時候，後海銀錠橋附近，就開了第一家酒吧。餘下的便是歷史。」

ChatGPT edit: 「約在2000年左右，當北京新興的夜生活圈子在舞場音樂退場後，想找個寧靜的地方休息，他們自行前往後海。差不多在這個時期，第一家酒吧在後海銀錠橋附近開業。從那時起，一個新的歷史故事就開始了。」 (... *Since that moment [sic], a new historical tale had begun [sic].*)

Once more, the paragraph generated is much longer, this time with three commas retained in one of the sentences. The meaning 'everybody knows the rest of the story' is again replaced by 'making history'.

This sentence, which distorted the author's intent, also introduced a grammatical error. It should be 新歷史故事在那時開始 ('a new historical tale began at that moment') instead of 新歷史故事從那時起開始 ('a new historical tale had begun since that moment').

Writing large numbers like Year 2000 in Chinese characters or Arabic numerals is solely a personal choice, so this is not a readability improvement. By adding 左右 ('about', 'approximately'), ChatGPT even created a new semantic overlap, as 左右 and 約 share the same definition (akin to 'about Year 2000 in approximate' in bad English). This makes the paragraph long and awkward, thus even less readable.

4 Inability to Correct Redundancy

Redundancy impairs readability by making the main point obscure.

4.1 Example - Threefold Redundancy

「除此之外，其他一切異點，都不必提起。」

ChatGPT edit: 「除此之外，不需要再提及任何其他異點。」

This is a simple example of threefold redundancy. 除此之外 is already a redundant blend of 除卻 and 此外, as they share the common definition – 'besides'. 其他 ('others') has the same meaning as well. Surprisingly, these redundant expressions are all retained in the 'revised' paragraph generated by ChatGPT.

4.2 Example - Paragraph with Both Redundancy and an Unnecessary Calque

Here is another sample of redundancy:

「大陸幾十個城市連續三天有反日示威，當局恐局勢失控，在各地採取拘捕行動。」

ChatGPT edit: 語義模糊：句子中的「當局恐局勢失控」的表達方式較為模糊，可能需要更清晰的表達。改善後的句子可能如下所示：「大陸幾十個城市連續三天發生反日示威，當局擔心局勢可能失控，於各地採取拘捕行動。」

ChatGPT's edit is totally mind-boggling. 「採取拘捕行動」 is an awkward fusion of 拘捕 ('arrest') and the calque 採取行動 from English 'take action'. The clause is redundant as 'arrest' and 'take action' refer to the same act. Besides, the author can opt for just 行動 as a verb, which already conveys the meaning.

Thus, 於各地採取拘捕行動 is expected to be changed to something like 在各地抓人. Yet, it is curious that these redundant terms remain unchanged in ChatGPT's response.

Inexplicably, the succinct terms 有 and 恐 were replaced by the lengthier 發生 and 擔心……可能. Generating a new answer only replaces 採取 with 展開, and the issue of redundancy remains.

Similar paragraphs yielded similar results, with the 'corrected' paragraphs becoming even less readable. 採取拘捕行動 remains in the paragraph, and the juxtaposition of two verbs in 破案拘捕 is inexplicable.

「遊客違反逗留條件，從事帶貨工作，警方會採取拘捕行動。」

ChatGPT edit: 「如果遊客違反逗留條件，從事非法帶貨活動，警方將採取拘捕行動。」

「兩岸與六個國家警察部門聯手，搗破一個跨境電信詐騙犯罪集團，於二十三日採取拘捕行動。」

ChatGPT edit: 「兩岸警方與六個國家的警察部門合作，成功破案拘捕了一個跨境電信詐騙犯罪團夥，行動在二十三日採取。」

5 Inability to Correct Confusing Format Switches

5.1 Example - Use of Arabic Numerals without Achieving Simplification

Arabic numerals are sometimes preferred over Chinese characters, that is when they can achieve simplification. Stock prices for instance, using \$12345.67 is much clearer than 一萬二千三百四十五元六毫七仙.

Yet, analogous to writing '1' and '2' instead of 'one' and 'two' in English, expressing small numbers as '1' and '2' instead of 一 and 二 in Chinese can lower the ease of reading. Writing place names like 九龍塘 as 9龍塘, and 沙田第一城 as 沙田第1城 in the following example (ChatGPT was only able to correct the former), is arguably even more ridiculous.

Question: 請找出以下句子可讀度低的地方。「1月2日上午3時，海關在沙田第1城5期某單位檢獲1公斤大麻，市值約80萬元。」

ChatGPT edit: 2023年1月2日上午03:00，海關在沙田第1城5期的某單位執行檢查，檢獲1公斤價值約80萬元的大麻。

All Arabic numerals are retained in ChatGPT's edit. ChatGPT even overspecified the time as 03:00 in a foreign format, which is baffling. 「一月二日上午三時，海關在沙田第一城五期某單位檢獲一公斤大麻，市值約八十萬元。」 is the most natural revision.

Other samples not only illustrate the bot's lack of understanding of fundamental rules, but also its Euro-centric nature. The original verse by Lu You (1125 – 1210) is 「三萬里河東入海，五千仞岳上摩天」 [1].

Question: 請找出以下句子可讀度低的地方。「30,000里河東入海，5,000仞岳上摩天」

ChatGPT edit: 以下是句子「30,000里河東入海，5,000仞岳上摩天」的可讀度低的地方：使用不常見的度量單位：這句話使用了不太常見的度量單位，如「里」和「仞」，這可能使讀者難以理解句子中的距離和高度。缺乏上下文：句子缺乏足夠的上下文，無法確定它所描述的具體場景或情境。為了提高這句話的可讀性，可以考慮轉換為更常見的度量單位，並提供一些上下文，例如：

「這條河流長達30,000公里，它流經河東地區最終注入大海；而這座岳山高達5,000米，幾乎觸及了天際。」

Question: 請找出以下句子可讀度低的地方。「30,000里河東入海，五千仞岳上摩天」

ChatGPT edit: 以下是句子「30,000里河東入海，五千仞岳上摩天」的可讀度低的地方：使用中文數字表示：句子中使用了中文數字「五千」，而前面使用了阿拉伯數字「30,000」，這種混用可能會讓讀者感到混淆。缺少標點符號：句子中缺少標點符號，導致兩個狀況（河流的長度和岳山的高度）之間的關係不夠清晰。要提高這句話的可讀性，可以考慮以下修正：

”30,000里的河東入海，以及5,000仞高的岳山摩天入雲。”

6 'False Positives' which Lead to Semantic Losses

6.1 Example - Change of Objects

In our analysis of GPT's edits, semantic losses are as prevalent as semantic shifts. These occurrences are observed across various parts of speech, including verbs, adjectives, and nouns.

「各地駐軍一定要密切配合地方黨委和政府，積極投身抗旱救災鬥爭，為打贏這場硬仗做出更大貢獻。」

ChatGPT edit: 「各地駐軍應緊密協作，積極參與抗旱和救災工作，以在這場嚴峻挑戰中發揮更大作用。」

In the new sentence, both the party committee and government (grammatical *objects*) were removed from the sentence. The object became the same as the subject - the troops, and the verb became *reflexive*. The idea of 'obeying the local party committee and government's orders' is lost, as the local party committee and government are not part of the army.

The actual issue is that there are too many stock expletives. A suggested readability improvement should be '各地駐軍必須聽從地方黨委和政府指揮，盡力抗旱救災。' [1].

6.2 Example - Meaning Loss & Meaning Shift

Here is another sample of ChatGPT's false positives:

「佳佳是香港接地氣的街坊超市，在荃灣的這家分店主要供應新鮮的肉類。」

ChatGPT edit: 「佳佳是香港一家位於荃灣的超市，專營新鮮肉類。」

It is observed that when ChatGPT encounters expressions it is uncertain or unfamiliar with, it tends to omit them in the edit. The descriptions of the supermarket being 'practical' (接地氣) and 'accessible' (街坊) were removed. Besides, there was a shift in meaning. From the original sentence, it is clear that the branch (分店) located in Tseun Wan (a district of Hong Kong) is not the supermarket's only site, yet ChatGPT's edit suggests the opposite. This indicates that ChatGPT may have a hallucination issue when handling Chinese texts.

7 Suggested Solutions

The GPT-4 model was trained with the help of *reinforcement learning from human feedback* (RLHF) on their private and proprietary evaluation platform. However, potentially due to the scarcity of experts knowledgeable about the nuances of foreign languages, the model may not perform as expected in certain language-specific tasks deemed straightforward.

Other factors why ChatGPT is ineffective at improving Chinese readability may be that the control of Chinese training data sources is not rigorous enough. It is intriguing that the source of our dataset, the online version of an educational column, is a widely recognised and publicly accessible website that ChatGPT can scrape and acquire knowledge from. Yet, ChatGPT may have missed the mark.

As publications like *Nature* and *Taipei Times* have pointed out, it is no surprise that places like Japan [7] and Taiwan [8] are making their own 'GPTs'. Many researchers have voiced that artificial intelligence systems trained on a foreign soil may not grasp the intricacies of their languages and cultures [7].

8 Conclusion

ChatGPT has a tendency to make unreadable written Chinese text even less readable, by unnecessarily lengthening them, introducing redundancies, removing or changing key terms, and so on. This raises concerns about the Chinese intention detection ability of the chatbot. We suspect that the issue is attributed to a lack of Chinese language experts involved in *reinforcement learning from human feedback* (RLHF), or inadequate control of Chinese training data sources.

References

- [1] Koo Tak-ming. 中華正聲. am730, 2014. Available: https://archive.am730.com.hk/column_index-162, accessed Sep. 25, 2023.
- [2] Rajendran Sankaravelayuthan. Lexical Gaps and Untranslatability in Translation. *Language in India*, 20(5):56–82, 2020.
- [3] Cambridge English Dictionary. Available: <https://dictionary.cambridge.org/dictionary/english/diplomatic>, accessed Sep. 25, 2023.
- [4] Longman Dictionary of Contemporary English. Available: <https://www.ldoceonline.com/dictionary/diplomatic>, accessed Sep. 25, 2023.
- [5] g0v Taiwan. Moedict (萌典). Available: <https://www.moedict.tw/%E6%89%8B%E6%AE%B5>, accessed Sep. 25, 2023.
- [6] Henry Watson Fowler. Vocabulary § Foreign Words. In *The King’s English (2nd ed.)*. New York, 1999.
- [7] Tim Hornyak. Why Japan Is Building Its Own Version of ChatGPT. *Nature*, September 2023. Available: <https://www.scientificamerican.com/article/why-japan-is-building-its-own-version-of-chatgpt>, accessed Sep. 25, 2023.
- [8] Yang Yuan-ting and Shelley Shan. Taiwan to develop its own ChatGPT. *Taipei Times*, February 2023. Available: <https://www.taipeitimes.com/News/front/archives/2023/02/14/2003794307>, accessed Sep. 25, 2023.