

# Concept-Centric Prompts and Translation Memory for Enhanced Machine Translation with Large Language Models

Ming Qian<sup>1</sup>, Chuiqing Kong<sup>2</sup>

<sup>1</sup>Charles River Analytics

<sup>2</sup>Freelance Translator

[mqian@cra.com](mailto:mqian@cra.com), [chuiqingkong17@gmail.com](mailto:chuiqingkong17@gmail.com)

## Extended Abstract

In cognitive science, a concept is perceived as a mental representation playing a pivotal role in diverse thought processes, thereby being essential in comprehending the ways the human mind processes information and embodies knowledge. We experimented with concept-driven GPT prompting to enable interactive machine translation (MT) learning/enhancement and incremental/ continual translation knowledge acquisition. We also propose to build a concept-centric machine translation memory system to support instance-based learning (by inferring the conceptual similarities between new and stored instances), continual learning (by continually collecting and refining translation-related concepts and the corresponding text instances), and concept formation to support customized preferences.

### Concept-driven prompts enhance machine translation revisions to achieve human-like quality

(Qian et al. 2023, 2024) experimented multiple types of translation instructions include pre-editing analysis, human-generated reference translation, and real-time translation revision based on human feedback centered around concepts or mental models (Table 1).

In (Qian et al. 2023), using multiple text segments, two professional human translators evaluate translation results generated by Google Translate (representing neural MT), a certificated human translator, GPT4 augmented MT using pre-editing analysis, GPT4 augmented MT using human-generated reference translation (see prompt examples in Table 1 and color scaled Likert rating scores in Figure 1). The results showed that the human translator outperformed all the other approaches, and GPT4 augmented MT methods showed superior results compared with the NMT results. The human translation reference approach, where the human expert’s translation output was added into the prompt instruction as additional reference, was the most interesting: while it is impractical to have human-generated translation for each sentence as part of the prompt, the performance of human + GPT4 results, on average, is actually worse than human-only result. These results led us to think that there should be a better way of human-machine collaboration to generate results that match human expert performance better.

Based on the belief that human and GPT communicate better using concepts and mental model descriptions, (Qian et al. 2024) experimented a new approach of real-time translation revision based on human feed-

back instruction centered around concepts or mental models. Table 2 shows two examples of inter-cultural mental model differences: most Chinese K-12 schools do not have school mascot, and most Chinese parents are not familiar with the concept of “travel-soccer chauffeurs”. After explaining the concepts in the instructions, GPT4 generated satisfactory outcome by making proper modifications. We also experimented with a variety of other concepts such as emphasizing aspect of necessity, focusing on ‘where’ instead of ‘how’, culture-specific idioms, different lexical choice, problematic omission, shorten or lengthen an expression, and preference to use metric units instead of English units. We found that concept-driven prompts lead to effective translation revision and achieve human-like translation quality (see color scaled Likert rating scores in Figure 2 based on evaluations by 2 professional translators). That leads to the belief that it is more effective for humans to explain the concepts than to provide human-generated translation text.

### LLM (GPT4) has knowledge of core translation concepts known to translation experts

To further validate our approach of using concept-driven prompt for MT revision, we focused on ten core concepts specific to Chinese-to-English translation:

1. **The concept of "pro-drop"** refers to the ability of a language to omit subjects (and sometimes other pronouns) when they are pragmatically or grammatically inferable. Chinese is a pro-drop language, meaning that subjects can be and often are omitted when they are understood from context. In contrast, English is a "non pro-drop" language, meaning that subjects are almost always explicitly stated, as English need the presence of subjects for clarity and grammatical correctness.
2. **Linguistic term dislocation** refers to the phenomenon where the positioning of words or phrases in a sentence differs significantly between two languages. This can pose challenges in translation, particularly between languages with distinct syntactic structures, like Chinese and English.
3. **The concept of implicit conjunctions and linking/transition words:** Chinese often omits conjunctions that would be necessary in English. This is because Chinese relies heavily on context and the inherent logic of the sentence structure. For instance, two clauses might be placed side by side without a conjunction, yet a native Chinese speaker would understand their relationship based on the context.

4. **The concept that English sentences tend to be longer while Chinese sentences are typically shorter** has implications for translation. English allows for complex sentences with multiple clauses, often linked by conjunctions, which can express detailed nuances and conditions. Chinese, on the other hand, favors brevity and conciseness.
5. **Deverbalization** is a concept in translation studies, particularly relevant in the context of translating between languages with significantly different structures, such as Chinese to English. The core idea of deverbalization is to move away from a direct, word-for-word translation, which often leads to awkward or inaccurate results due to the differences in grammar, syntax, and cultural contexts. Instead, the translator interprets the meaning of the source text at a deeper, more abstract level, and then re-expresses this meaning in the target language.
6. **Repositioning syntax components** in the context of Chinese-to-English translation refers to adjusting the order and structure of words and phrases to ensure that the translated text conforms to the grammatical and stylistic norms of English, while maintaining the original meaning and intent of the Chinese text. This process is necessary because Chinese and English have different syntactic structures.
7. **Changing subject selection** involves selecting a subject different from the one in the source text to improve the readability and fluency of the English translation.
8. **Changing the perspective** in the context of Chinese-to-English translation refers to adapting the viewpoint or approach when translating text from Chinese into English. This concept is crucial due to significant differences between the two languages in terms of grammar, syntax, idiomatic expressions, and cultural contexts.
9. **Changing parts of speech** is to faithfully convey the meaning of the original at the same time accords with English means of expression irrespective of the part of speech of a word in Chinese.
10. **Cultural translation** aims to show the cultural differences while respecting the source culture as well.

For all the examples we experimented with, the initial translation results are unsatisfactory. Firstly, we queried the GPT4 about the concept and verified that the GPT4 knows about the concept. Then we requested the GPT4 to revise the translation and all revised results had the proper correction related to the concept. Table 2 shows an example related to the concept of “the pro-drop nature of Chinese and the non pro-drop nature of English”.

#### **Enabling Human-centered Machine Translation with Concept-driven Large Language Model Prompts and Concept-centric Machine Translation Memory**

Commercial Translation Memory (TM) systems have been used for over 20+ years (Reinke 2018). TM stores the original source text and the translated target text in a computer system so that translators can recycle these translated segments by automatically proposing a relevant translation from the memory as a complete (“exact match”) or partial solution (“fuzzy match”) whenever the same or a similar sentence occurs again.

Based on our studies presented in the previous section, we propose to build a concept-centric translation memory to recycle concepts instead of text. Such a system can suggest a relevant translation as an exact or fuzzy concept match whenever sentences associated with a similar concept occur again. Since LLMs were found to be very effective—in terms of matching human judgment—to measure semantic concept similarity (Le Mens et al. 2023), we used GPT-4 to derive similarity scores (between 0 and 100) of text segments to translation concepts and compare the scores with human experts’ ratings. The empirical test results show that the GPT-4 derived scores match well with the human generated evaluations.

Such a system is an instance-based learning system that supports learning by incrementally storing training instances in memory and using them to make concept predictions. When a new text instance is encountered, the algorithm finds the closest stored instances based on concept similarity measures, and then generates the corresponding translation using existing training instances corresponding to the same concept. Table 4 shows an example of predicting a new instance on the pro-drop versus non pro-drop concepts based on the concept definition and 2 training instances. We collected 3 examples each for the ten Chinese-to-English translation concepts described above and performed a leave-one-out 3-folder validation. The high classification accuracy indicates the feasibility of a concept-centric translation memory approach where the learning is human-like—interactive, incremental, continual, and few-shot.

In addition, various strategies (e.g. providing free-form text description, few-shot examples, or a combination of both) can be employed for translators to define customized concepts to reflect their own preferences and styles.

#### **Conclusion**

The proposed concept-driven GPT prompts and concept-centric machine translation memory system offer several advantages:

- They create novel MT systems that exhibit human-like learning characteristics: interactive, incremental, continual, rapid knowledge acquisition based on few-shot examples without costly re-training,
- They humanize MT by enabling interactions that mimic human concept-driven collaborated understanding.
- Unlike traditional neural MT approaches, often perceived as opaque 'black boxes' by human translators, this new method fosters a collaborative environment. It does so by emphasizing concept description and prediction, bridging the gap between human and machine.
- This approach also addresses the issue of human translators feeling undervalued by technology. By working through initial examples, the translation memory system can automate future tasks related to the learned concepts. This allows machines to handle the monotonous and repetitive aspects, freeing human translators to tackle increasingly challenging, nuanced, and complex problems.

## References

- Le Mens, G., Kovács, B., Hannan, M.T. and Pros, G., 2023. Uncovering the semantics of concepts using GPT-4 and Other recent large language models (No. 1864).
- O'Brien, S., 2012. Translation as human-computer interaction. *Translation spaces*, 1(1), pp.101-122.
- Qian, M., et al. 2023, September. Performance Evaluation on Human-Machine Teaming Augmented Machine Translation Enabled by GPT-4. In *Proceedings of the First Workshop on NLP Tools and Resources for Translation and Interpreting Applications* (pp. 20-31).
- Qian, M., et al. 2024, May. Augmented Machine Translation Enabled by Generative AI Large Language Models: Performance Evaluation on Human-Machine Teaming Approaches. In *Proceedings of the 4th IEEE International Conference on Human-Machine Systems* (submitted).
- Qian, M., et al. 2024, July. Enabling Human-centered Machine Translation with Concept-driven Large Language Model Prompting and Machine Translation Memory. 5th International Conference on Artificial Intelligence in HCI (AI-HCI), Part of HCI International Conference, Washington DC, USA (accepted).
- Reinke, U., 2018. State of the art in translation memory technology. *Language Technologies for a Multilingual Europe*; Rehm, G., Stein, D., Sasaki, F., Witt, A., Eds, pp.55-84.

GPT augmented MT approaches	Prompt examples
Pre-editing analysis	Please read the story at <a href="https://www.thecut.com/2020/03/book-excerpt-samantha-irbys-wow-no-thank-you.html">https://www.thecut.com/2020/03/book-excerpt-samantha-irbys-wow-no-thank-you.html</a> and make a list of the author's writing styles. Translate the following English sentence to Chinese reflecting the writing styles.  Original English: ...  The Chinese translation is: ... (GPT4 output)
GPT4 augmented MT using human-generated reference translation	Translate the following text segment from English to Chinese:  Original English: ...  Chinese translation (by the human translator) is: ...  The Chinese translation is: ... (GPT4 output)
GPT4 augmented MT using human guidance on concepts and	Translate the following text segment from English to Chinese:  Original English: ...

mental models	The Chinese translation is: ... (GPT4 output) Revise the translation based on: ... (Human guidance on concepts and mental models)  The Revised Chinese translation is: ... (GPT4 revised output)
---------------	--

Table 1: Prompt examples for different types of GPT4 augmented MT approaches (Qian et al. 2023 and Qian et al. 2024).

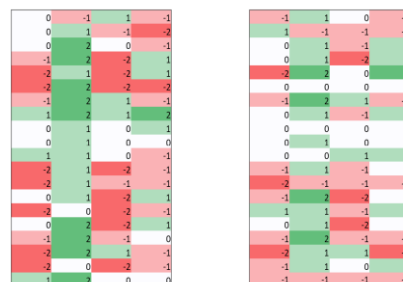


Figure 1: Color scaled Likert rating scores (5-point range from -2 to 2) by the two evaluators (Qian et al. 2023). Four columns are for Google Translate representing NMT, a human expert, GPT4-LLM augmented MT using style analysis as instruction, and GPT4-LLM augmented MT using human-generated translation reference as instruction.

Source Text (English)	Original Translation	Concept-driven Instruction	Modified Translation
The last time I wore a Halloween costume was in the second-grade costume parade at Lincoln Elementary School in 1986 (go, dolphins).	我最后一次穿万圣节服装是在1986年林肯小学二年级的服装游行上（加油，海豚）。	In China, most schools do not have mascot. Revise the translation to make it clear to Chinese audience that dolphin is a school mascot.	我最后一次穿万圣节服装是在1986年林肯小学二年级的服装游行上，那时我们学校的吉祥物是海豚（加油，海豚）。
A few years ago (before I moved to Michigan and joined my wife's community of backyard composters and travel-soccer chauffeurs), my lady and I went to her friend's	几年前（在我搬到密歇根州并加入我妻子的后院堆肥社区和旅行足球司机圈子之前），我和我的女士参加了她朋友的化妆婚礼。	Revise the translation by elaborating more on the concept of "travel-soccer chauffeurs" since Chinese audience might not be familiar with this concept.	几年前（在我搬到密歇根州并加入我妻子所在的那个社区，那里的居民们不仅会在后院进行堆肥，还会担任孩子们“巡回足球”的司机，负责接送他们参加

costume wedding.			各地的足球比赛), 我和我的伴侣参加了她朋友的主题化妆婚礼。
------------------	--	--	--------------------------------

Table 2: two examples of GPT4 augmented real-time MT translation revision employing human guidance on concepts and mental models.

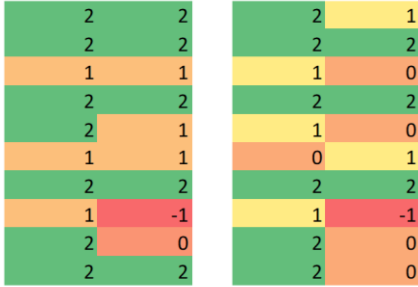


Figure 2: Color scaled Likert rating scores (5-point range from -2 to 2) by the two evaluators (two columns represents the two evaluation questions: (1) Does the revised translation follows the concept-driven human instruction provided in the prompt? (2) Does the revised translation achieve the same accuracy as the human-generated translation in terms of the concept described in the instruction? The results show that the concept-driven instructions lead to related revision and achieve good human-like performance.

**User:**

Translate the following Chinese segment into English:

Chinese: 深刻理解疫情的经济影响, 十倍努力推动经济反弹。

English translation:

**ChatGPT:**

The English translation is:

"Deeply understand the economic impact of the pandemic, and make tenfold efforts to promote economic rebound."

**User:**

Are you familiar with the concept of "pro-drop" and "non pro-drop" language? Especially in the context of Chinese-to-English translation?

**ChatGPT:**

Yes, I am familiar with the concept of "pro-drop" and "non pro-drop" languages, especially in the context of translation between Chinese and English.

The term "pro-drop" refers to languages in which certain pronouns (usually subjects) can be omitted when they are pragmatically or grammatically inferable...

Chinese is a typical pro-drop language. In Chinese, subjects, objects, and even some other elements of a sentence can be omitted if they are understood from the context. English, on the other hand, is a "non pro-drop" language...

In English, subject pronouns are usually mandatory, and dropping them often leads to sentences that sound incomplete or grammatically incorrect...

The pro-drop characteristic of Chinese often leads to more concise sentences, whereas English usually requires more explicit expression of subjects and sometimes objects. Thus, when translating from Chinese to English, it's important to not only translate the words but also interpret and convey the implied subjects or objects to make the sentence comprehensible in English.

**User:**

Alright, can you revise the translation of the Chinese segment "深刻理解疫情的经济影响, 十倍努力推动经济反弹。" based on the concept of "pro-drop" and "non pro-drop" languages.

**ChatGPT:**

Absolutely, considering the pro-drop nature of Chinese and the non pro-drop nature of English, a revised translation of the Chinese segment "深刻理解疫情的经济影响, 十倍努力推动经济反弹。" could be:

"Deeply understanding the economic impact of the pandemic, we must exert tenfold efforts to drive an economic rebound."

Table 3: In this example related to the concept of "the pro-drop nature of Chinese and the non pro-drop nature of English", initial translation result is unsatisfactory and we query the GPT4 about the concept and it turns out the GPT4 knows about the concept. Then we request the GPT4 to revise the translation and the revised result reflects the proper correction related to the concept.

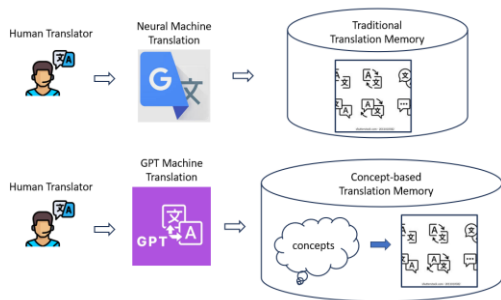


Figure 3: Traditional Translation Memory Versus Concept based Translation Memory.

**User:**

You are an expert-level Chinese-to-English translator. Core Chinese-to-English translation concept: The concept of pro-drop refers to the ability of a language to omit subjects (and sometimes other pro-nouns) when they are pragmatically or grammatically inferable. Chinese is a pro-drop language, meaning that subjects can be and often are omitted when they are understood from context. In contrast, English is a non pro-drop language, meaning that subjects are almost always explicitly stated, as English needs the presence of subjects for clarity and grammatical correctness.

The core translation technique to handle the concept is to use passive voice.

**Example 1:**

**Chinese:** 要重视发挥家庭、学校和社区作用，帮助儿童抵御恐怖极端思想侵蚀。

**English:** Attention should be paid to the roles of families, schools, and communities in preventing children from being influenced by extreme terrorism.

**Example 2:**

**Chinese:** 对传统那些没有比较优势和竞争优势产业，继续给予必要保证。

**English:** Traditional industries without comparative advantage and competitiveness should be given the support. Here is another Chinese text segment: 放开传统受到抑制，具有一定比较优势、劳动密集型产业，放开给乡镇企业、私营企业和外资企业来进入。

How typical is the Chinese text segment related to the core translation concept and examples described above? Provide your response as a score between 0 and 100 where 0 means 'not typical at all' and 100 means 'Extremely typical'. If the score is larger than 75, generate the English translation accordingly.

**ChatGPT:**

The Chinese text segment "放开传统受到抑制，具有一定比较优势、劳动密集型产业，放开给乡镇企业、私营企业和外资企业来进入。" is highly typical in relation to the core translation concept of "pro-drop"; and the examples

provided. In this segment, the subject is not explicitly mentioned, which is a common characteristic in Chinese due to its pro-drop nature. The context implies the subject, which is a common practice in Chinese language but less so in English. Based on these observations, I would assign **a score of 90**, indicating it is extremely typical of the pro-drop feature in Chinese. Given the score is larger than 75, here is the English translation following the core translation technique of using passive voice:

"Restrictions on traditional industries that have been suppressed but possess certain comparative advantages, particularly labor-intensive ones, should be lifted to allow for the entry of township enterprises, private companies, and foreign-funded enterprises.";

Table 4: GPT-4 based concept description and few-shot training examples can predict a new instance accurately and generate the corresponding translation effectively.