

# Exploring the Gap: The Challenge of Achieving Human-like Generalization for Concept-based Translation Instruction Using Large Language Models

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

<sup>1</sup>Charles River Analytics

<sup>2</sup>Independent Researcher

mqian@cra.com, chuiqingkong17@gmail.com

## Abstract

Our study utilizes concept description instructions and few-shot learning examples to examine the effectiveness of a large language model (GPT-4) in generating Chinese-to-English translations that embody related translation concepts. We discovered that human language experts possess superior abductive reasoning skills compared to GPT-4. Therefore, it is crucial for humans to employ abductive reasoning to craft more detailed instructions and infuse additional logic into exemplary prompts, a step essential for guiding a large language model effectively, in contrast to the more intuitive understanding a human expert might have. This approach would make the prompt engineering process more complicated and less human-like. Emphasizing domain-specific abductive reasoning stands out as a crucial aspect of human-like learning that AI/ML systems based on large language models should aim to replicate.

## Concept-based Machine Translation Workflow Enhanced by Large Language Models

John McCarthy's insight, "To understand natural language is to understand the concepts in the language, not just the words," underscores the essential role of grasping the concepts conveyed in texts. Recent advancements in generative Artificial Intelligence (GenAI), particularly through large language models (LLMs) like GPT-4, have demonstrated their effectiveness as machine translation tools in various studies and evaluations (Jiao, Wang, and Huang 2023; LinkedIn Pulse 2023). Unlike traditional machine translation (MT) tools such as Google Translate and DeepL, which rely solely on the source language text, GPT-4 possesses the ability to follow instructions and can also learn from bilingual examples provided within the prompt. This capability enables GPT-4 to produce translations that are more contextually aware and accurate, positioning it as a more flexible and user-friendly alternative to conventional methods.

## Challenge of Achieving Human-like Generalization for Concept-based Translation Instruction Using Large Language Models

The concept of generalization in AI involves the ability of a model to perform well on new, unseen instances. Human-like generalization in natural language processing (NLP) refers to the ability of NLP models to generalize in a way that is similar to how humans do.

In our research (Qian, et al., 2023; Qian and Kong, 2024), we explored how effectively GPT-4 can clarify translation concepts, evaluate their significance in source texts, and accurately convey them in the target language, specifically through examples from Chinese to English. Our results show that by employing instructions for concept description along with few-shot learning examples, GPT-4 successfully elucidates most concepts, precisely assesses their relevance, and translates them into the target texts, embodying the related concepts. Nonetheless, achieving satisfactory performance for a considerable array of concepts necessitates the use of advanced prompting techniques, such as the Chain-of-Thought (CoT), or pre-editing approaches, such as explicating linguistic patterns. Therefore, despite their advancements, LLMs like GPT-4 still exhibit limitations in terms of their human-like generalization capabilities for concept-based translation tasks.

To illustrate, we utilized a straightforward concept definition as a part of the prompt for GPT-4: "Changing subject selection involves selecting a different subject than the one in the source text to enhance the readability and fluency of the English translation." Additionally, four instances related to the subject changing selection concept were provided as part of the prompt. Every sentence has a noun phrase composed of two nouns, where one noun modifies the other.

1. 资源环境约束边界临近, 最典型的例子就是雾霾。(direct translation: **The boundaries of resource and environmental constraints are approaching. The most typical example is smog.**)

2. 老年人本身的免疫功能就相对于年轻人要弱一些。(direction translation: **The immune function of the elderly is weaker than that of the young.**)
3. 未来20年气候变化的威胁将超过恐怖主义。(direct translation: **The threat from Climate change will surpass terrorism in the next 20 years.**)
4. 迁移的流向和形式也都发生了很大的变化。(direct translation: **The flow and form of migration have also undergone great changes.**)

We noted that examples #2, #3, and #4 utilized the possessive particle "的" to connect two nouns, signifying possession or a relationship and necessitating a shift in the subject from one noun to another for the translation into the target language (English). In contrast, example #1 did not use the possessive particle "的" to link two nouns, but its existence was implied. This variation could account for the inadequate 1-in-4 leave-one-out cross-validation (LOOCV) performance. To make the task more straightforward for GPT-4, we replaced example #1 with another example in which the possessive particle was explicitly used.

1. 技术的进步极大地改变了我们的生活方式。(direct translation: **The advancement of technology has greatly changed our way of life.**)

Subsequently, we checked the 1-in-4 leave-one-out cross-validation (LOOCV) performance again. However, the outcomes were not satisfactory either. On several occasions, GPT4 did not utilize the two nouns linked by “的” but instead selected other nouns within the sentence. This necessitated our provision of comprehensive, step-by-step analyses for the few-shot examples included in the prompt, as illustrated in Table 1.

The performance of the 1-in-4 leave-one-out cross-validation (LOOCV) was significantly enhanced by the detailed reasoning provided through the Chain-of-Thought (CoT) method. Nonetheless, its practical application is constrained by the necessity to deconstruct the original concept into a detailed, step-by-step reasoning process.

Unlike human translators, who can generalize a concept across varied contexts by recognizing both explicit and implicit patterns and focusing on the core intent, the GPT-4-based instruction plus few-shot approach relies on identifying explicit patterns through detailed descriptions. For reasons of space, the CoT strategies for the other translation concepts are not detailed here.

## Discussion

Abduction is an inference process that identifies the most plausible explanations for observed phenomena. Human language experts exhibit superior abduction skills compared to GPT-4. For instance:

- (1) They understand that the application of the translation concept is not affected by whether explicit or implicit particles are used to link two nouns.

- (2) They can recognize, as shown in the few-shot examples, that when changing the subject in a sentence that starts with 'A's B,' they need to select either A or B without choosing other nouns in the sentence.

In comparison, GPT-4 struggles with such abductive inference reasoning.

As a result, human experts must engage in abduction to derive specific instructions (deductive rules) and develop step-by-step logic in example prompts (inference demonstration). This necessity complicates and extends the prompt engineering process. Emphasizing domain-specific abductive reasoning stands out as a crucial aspect of human-like learning that AI/ML systems based on large language models should aim to replicate.

<p>迁移的流向和形式也都发生了很大的变化。  <b>Direct translation:</b> The direction and form of migration have also undergone significant changes.  <b>The subject phrase:</b> The direction and form of migration  <b>Translation with alternative subject selection using the other noun in the subject phrase:</b> Migration was happening in different directions and ways.</p>
<p>技术的进步极大地改变了我们的生活方式。  <b>Direct translation:</b> The progress of technology has greatly changed our way of life.  <b>The subject phrase:</b> The progress of technology  <b>Translation with alternative subject selection using the other noun in the subject phrase:</b> Technology have greatly advanced to change our way of life.</p>
<p>老年人本身的免疫功能就相对于年轻人要弱一些。  <b>Direct translation:</b> The immune function of elderly people is relatively weaker than that of younger people.  <b>The subject phrase:</b> The immune function of elderly people  <b>Translation with alternative subject selection using the other noun in the subject phrase:</b> Unlike younger people, the elderly have weaker immune systems.</p>
<p>未来20年气候变化的威胁将超过恐怖主义。  <b>Direct translation:</b> The threat of climate change will exceed that of terrorism in the next 20 years.  <b>The subject phrase:</b> The threat of climate change  <b>Translation with alternative subject selection using the other noun in the subject phrase:</b> In the next two decades, climate change will pose a greater threat than terrorism.</p>

Table 1: Four instances related to the concept of changing subject selection, demonstrating how Chain-of-Thought (CoT) prompting facilitates reasoning through smaller, more manageable steps.

## References

- Jiao, W., Wang, W., Huang, J.T., Wang, X. and Tu, Z.P., 2023. Is ChatGPT a good translator? Yes with GPT-4 as the engine. arXiv preprint arXiv:2301.08745.
- LinkedIn Pulse. 2023. How GPT-4 Is Transforming the Language Service Industry: Benefits and Challenges. <https://www.linkedin.com/pulse/how-gpt-4-transforming-language-service-industry-benefits-kotzsch/>. Accessed: 2024-02-07.
- Qian, M., Wu HQ., Yang, L., Wan A., 2023. 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, pages 20–31, Varna, Bulgaria.
- Qian, M.; and Kong, CQ. 2024. Enabling Human-centered Machine Translation Using Concept-based Large Language Model Prompting and Machine Translation Memory. Paper has been accepted and will appear In International Conference on Human-Computer Interaction. Washington DC, USA, June 6<sup>th</sup> till July 4<sup>th</sup>.