

Can Artificial Intelligence Understand and Talk about Metaphors? :

An Empirical Study on the Metaphoric Competence of ChatGPT

Haoyu LU * Tao HAN **
Yutao SUN *** Chuan LIU ****

Abstract

This article introduced a metaphorical competence test scheme tailored for Artificial Intelligence language models and conducted a comprehensive evaluation of ChatGPT's metaphorical competence across three levels: metaphor identification, comprehension, and generation. The study revealed the following findings: (1) ChatGPT exhibits robust metaphor identification competence at the lexical level, but displays shortcomings in recognizing certain metaphors, such as container metaphors and primary metaphors. (2) While ChatGPT excels in understanding the conventionality of language expression and metaphorical inference tasks, it may face challenges in tasks related to comprehending similarity and inducing various metaphor types. (3) Regarding metaphor generation, ChatGPT

* Haoyu LU, Kyushu International University, h-lu@cb.kiu.ac.jp

** Tao HAN, Beijing Foreign Studies University, hantao@bfsu.edu.cn

*** Yutao SUN, Beijing Foreign Studies University, songyokutou@bfsu.edu.cn

**** Chuan LIU, Beijing Foreign Studies University, 20010035@bfsu.edu.cn

tends to produce conventional metaphors and lacks sufficient novelty. The future direction of research will explore topics such as enhancing the language proficiency of artificial intelligence and conducting comparative studies between human and machine in terms of metaphorical competence.

KeyWords: language intelligence; metaphorical competence; metaphor identification; metaphor comprehension; metaphor generation

・ Introduction

Language intelligence is a topic of common concern in linguistics, cognitive science, and computer science (Zhou et al., 2017). Li and Liang (2022) pointed out that today, as artificial intelligence has developed to the stage of cognitive intelligence, “language intelligence” is an emerging interdisciplinary field that urgently needs to be established and developed. In particular, with the widespread application of cognitive linguistics and conceptual metaphor theory in the field of second language acquisition, metaphorical competence is seen as an important manifestation of language and communicative competence (Bailey, 2003; Danesi, 1986, 1994). Cognitive linguists believe that conceptual metaphors form the basis of much human thought, and its representation may occur at various levels of language such as words, sentences, and discourse. Second language acquisition researchers have found that the ability to produce and comprehend metaphor is essential in second language learning, as metaphor can be used to aid comprehension, extend thought, shed new light on old issues, and clarify ideas (Littlemore, 2001; Low, 1988).

Recently, the chatbot ChatGPT launched by the American company OpenAI has attracted widespread attention worldwide for its language

intelligence performance. Although world-renowned linguist Chomsky said in a February 2023 interview that ChatGPT is “basically high-tech plagiarism” and “a way of avoiding learning”, the concern arises precisely from ChatGPT’s knowledge base, which far exceeds that of ordinary people and its powerful language intelligence never seen in artificial intelligence before. So, as a high-intelligence chatbot, how is ChatGPT’s metaphorical competence? In order to answer this question, this article intends to draw on and improve previous test scheme on metaphorical competence to comprehensively test and evaluate ChatGPT’s metaphorical competence under English conditions.

• Background

• Conceptual Metaphor Theory

Conceptual metaphor theory suggests that human conceptual systems and abstract thinking are inherently metaphorical, and metaphorical expressions are the outward manifestation of the conceptual metaphors in the human mind (Lakoff & Johnson, 1980; Lakoff & Johnson, 1999). Metaphor is not only a rhetorical device used to decorate language, but also a cognitive mechanism. This theory defines metaphor as “understanding one conceptual domain in terms of another conceptual domain” (Kövecses, 2002, p.4). For example, we can use A JOURNEY to understand and talk about LOVE (such as “Look how far we’ve come.”). And, more importantly, we can also make metaphorical inferences on the target domain based on the relevant knowledge from the source domain, as in (1).

(1) Marriage is like an old car that you are accustomed to driving... If the car breaks down halfway, it’s a big problem for you. Do you stand by the

roadside and wait for it, or do you hitch a ride and continue on your journey, waiting for it to be towed back while it waits for you at home? Or do you simply give up on it and get a new car, but you can't reach your destination on foot. No matter how old you are, you still need a means of transportation. (Liu Liu, Wangtan Yu Fenghua)

・ Metaphorical Competence

So far, there have been many discussions on metaphorical competence. For example, Danesi (1993) defined metaphorical competence as the ability to understand and use metaphor in natural communication, and considered it a basic characteristic of native speakers' ability and the foundation of conceptual fluency. Littlemore (2001) described and measured four aspects of metaphoric competence: (a) originality of metaphor production, (b) ability to find meaning in metaphor, (c) speed in finding meaning in metaphor, (d) fluency of metaphor interpretation. Littlemore and Low (2006) further pointed out that metaphorical competence includes knowledge of, and ability to use metaphor, and skills needed to work effectively with metaphor.

Similarly, Chinese academic Zhong Yong (2015, p.53) divided the constituents of metaphorical competence into the ability to recognize, understand, and produce metaphorical expressions, and pointed out that the ability to understand is the basis of the ability to recognize and produce, and the development of the ability to recognize is based on the development of the ability to produce. In terms of practicality, the ability to understand and produce metaphorical expressions are more important than the ability to recognize.

Therefore, based on the definitions of metaphorical competence and the testing and evaluation methods by Littlemore (2001) and Zhong (2015),

the present paper will empirically examine the metaphorical competence of ChatGPT from three levels: metaphor identification, comprehension and generation.

• Methodology

• Research Object

The object of this study is the intelligent chatbot ChatGPT launched by the American company OpenAI (tested from March 26 to 28, 2023), using the OpenAI GPT-3 version. Its training data come from internet texts and other resources before 2021, including Wikipedia, news articles, social media, forums, and more. ChatGPT uses these data for large-scale self-supervised learning, enabling it to demonstrate excellent performance in multiple fields.

• Terminology Definition and Testing Plan

As previously mentioned, this article divides metaphorical competence into three levels: metaphor identification, comprehension, and generation. The definitions of relevant terms are as follows:

(a) Metaphor Identification: the ability to identify metaphorical expressions;

(b) Metaphor Comprehension: the ability to accurately understand metaphors, including the understanding of similarity, metaphor classification, and reasonable reasoning abilities;

(c) Metaphor Generation: the ability to generate reasonable and novel metaphorical expressions.

Prior to the formal testing, we conducted pre-testing using a small amount

of language data and topics (see Table 1) and optimized the questioning methods based on ChatGPT's responses.

Through pre-testing, we found that ChatGPT has a basic understanding of conceptual metaphor theory and is able to provide reasoning when identifying whether a given corpus contains conceptual metaphor. Based on this, this paper first designed a test plan for metaphor identification at the lexical and sentence level (see Table 2).

Secondly, this article designs four test plans for metaphor comprehension based on the following four indicators (see Table 3):

Finally, we test ChatGPT's ability to generate metaphors by using specified fixed concepts to allow it to freely generate poetry and dialogue text, and then manually evaluate the generated results (see Table 4).

• Results and Discussion

• Metaphor Identification Competence

• Lexical Level

The test materials at the lexical level include five words and five phrases (all containing metaphorical meanings) (see Table 5). If ChatGPT successfully provides one of them, it is considered a correct answer.

The test results show that ChatGPT answered “Yes” to whether the above words have a metaphorical meaning, and therefore achieved a 100% accuracy rate. However, for the phrase “the foot of a hill”, ChatGPT initially stated that it “typically has a literal meaning”, but later pointed out that it may have metaphorical meanings or usages depending on the context in which it is used. Thus, it was still considered to have passed the test.

• Sentence-level

At the sentence level, there are 6 sets of metaphoric identification test materials, with a total of 27 sentences (see Table 6). If the results given by ChatGPT are completely consistent with the existing research results, or if they are not completely consistent but reasonable, they will be considered as correct answers.

ChatGPT demonstrated a 100% accuracy rate by correctly identifying all five sentences in the first group as metaphorical expressions, accompanied by reasonable explanations. For example, ChatGPT's answer to (1a) was "Yes, the sentence 'Is that the foundation for your theory?' can be considered a metaphorical expression. In this context, 'foundation' is being used metaphorically to mean the basis or fundamental principle on which a theory is built."

For the second group of 7 sentences that inquired about the existence of metaphorical expressions, ChatGPT initially responded with a "No" for all except (2e), denying the presence of metaphorical meanings. However, after considering container metaphors as a standard and conducting a reassessment, ChatGPT concluded that, apart from (2a) and (2e), all the sentences do not contain container metaphors. The judgment of (2a) was in line with existing research, which regarded "race" as a container metaphor, while the judgment of (2e) deviated from existing research but had some validity ("running is a container that can hold different levels of skill or ability"). Therefore, out of the total 7 sentences, one passed the first test, resulting in a 14.29% accuracy rate, while two passed the second test, resulting in a 28.51% accuracy rate. The average accuracy rate for both tests was 21.40% (rounded to two decimal places).

For the third group of sentences, ChatGPT believed that it may not

necessarily be a metaphorical expression and could simply be a literal meaning that rocks become brittle with time. However, ChatGPT also provided a possible metaphorical meaning, which is to metaphorically represent the aging of a person or organization. Therefore, we then asked ChatGPT to re-evaluate, and the second answer clearly affirmed that the sentence is a metaphorical expression that metaphorically represents the aging of a person (“Yes, the sentence ‘The old rock is becoming brittle with age’ contains a metaphorical meaning… This is a metaphor for the human aging process, where we also become more fragile and vulnerable as we grow older.”). Therefore, the accuracy rate for this group of testing was 100%.

For the fourth group of 5 sentences inquired whether they are metaphorical expressions, ChatGPT answered “Yes” for all except (4b). For example, for (4a), ChatGPT’s answer was “Yes, the sentence ‘This building looks like a vast ocean’ contains a metaphorical meaning. The use of the word ‘ocean’ suggests that the building has certain qualities that are similar to those of an ocean.” However, for (4b), ChatGPT’s answer was “The sentence ‘Her long hair is like a pill’ does not make sense as a metaphor because there is no clear connection or similarity between long hair and a pill.” Therefore, out of the 5 sentences, only (4b) did not pass the test, resulting in an accuracy rate of 20%.

For the four sentences in Group 5, except for (5d), ChatGPT considered all other sentences to be metaphorical expressions and provided appropriate explanations. However, although ChatGPT denied that the sentence (5d) itself is a metaphor, it pointed out that “plumbing” is used metaphorically in the sentence to refer to the internal workings or components of a complex theory or system. Therefore, we considered it as a pass for the test. The accuracy rate for this group of tests was 100%.

With regards to the five sentences in Group 6, except for (6c) and (6d),

which were considered metaphorical expressions and appropriate judgments were given, ChatGPT did not accurately identify the other three sentences. For example, ChatGPT's response to (6a) was "The sentence 'They greeted me warmly' is not a metaphorical statement. It is a literal statement that means that the people being referred to welcomed or received the speaker with warmth or friendliness." Therefore, the final accuracy rate for this group of tests was 40%.

• Discussion

Overall, ChatGPT has strong metaphor identification competence, but there are also some issues.

Firstly, ChatGPT is not very sensitive in identifying certain relatively fixed metaphors, such as the container metaphor in Group 2 and the primary metaphor in Group 6. The reason may be that these metaphors are considered "dead metaphors", leading the machine to perceive them as fixed language chunks during the learning process, without actively discriminating the metaphors embedded within them. In particular, the test results in Group 6 showed that ChatGPT's metaphor identification ability for adjectives is weak (such as "Prices are high").

Secondly, ChatGPT demonstrates a robust ability to identify novel metaphors or expressions with semantic conflicts, displaying a keen sensitivity towards locating semantic contradictions. However, when assessing expressions that violate the consistency principle of metaphors, it tends to exhibit an inclination for excessive or one-sided interpretations, leading to distorted identification results, as observed in the results of Group 4. Therefore, ChatGPT still needs to strengthen its judgment of the rationality of metaphor expressions and the similarity evaluation involved in metaphors.

• Metaphor Comprehension Competence

The metaphor comprehension competence test is divided into four groups, each of which is aimed at different types of metaphors (see Table 7). If the results given by ChatGPT are consistent with previous research results or are reasonably consistent, they will be considered as correct answers.

• Conventional Judgments

According to Katz's (1996) research results, the novelty of the 10 sentences used in this test from strongest to weakest is as follows: Handshakes are doorbells. > Their love is a coniferous tree. > Their friends are houseplants. > This textbook is a winning lottery ticket. > His smiles are can openers. > His standards are at the ceiling. > That comment has the bull's eye. > She is a magnet to men. > That experience was a crushing blow. > Children are precious gems.

For the convenience of testing, we arranged the above sentences in a gradient and divided them into five groups with decreasing novelty difference between the sentences within each group, with the difficulty level increasing from Group A to Group E. The novelty of each comparison group is: (A1) > (A2); (B1) > (B2); (C1) > (C2); (D1) < (D2); (E1) < (E2). ChatGPT's results for the five comparison groups were consistent with the above results, and reasonable justifications were provided. Therefore, the accuracy of this test was 100%.

• Similarity Understanding

In this group, we provided a corpus and asked ChatGPT whether the sentence is reasonable or makes sense.

Regarding (1), Pinker (1994, p.295) explains, "It consumes a fifth of the

body's oxygen and similarly large portions of its calories and phospholipids. Greedy neural tissue lying around beyond its point of usefulness is a good candidate for the recycling bin.” However, ChatGPT considered (1) to be an unreasonable or meaningless expression because “it is not accurate or logical to equate the metabolic processes of the brain with those of a pig.” It can be seen that its processing of the sentence mainly compared the metabolism of the human brain with that of a pig, without linking the oxygen and energy consumption of the brain to that of a pig's food intake.

Regarding (2), Basso (1976, p.249) explains, “Ravens are widows, these people say. They say that because ravens and widows are poor and don't have anyone to get meat for them. This is why sometimes these women will stand near your camp and wait like that until you give them food. It is the same way with ravens. They stand around near the roads so they can eat what is killed there; they just wait like that until some car hits something and kills it. That's what ravens are doing when you see them standing near roads. They are waiting to get fresh meat.” However, ChatGPT considered (2) to be meaningless and seemingly unreasonable because “The sentence ‘Ravens are widows’ does not make sense or seem reasonable because it equates two unrelated things - ravens and widows - without any context or explanation. Ravens are a type of bird that belong to the crow family and are known for their black feathers and intelligence. Widows, on the other hand, are women whose spouses have died. There is no clear connection or comparison between these two things.” However, in reality, (2) is a special word order in the Apache language. The source domain of this metaphor is raven, and the target domain is widow. The Apache people's understanding of this metaphor relies on their knowledge of the features of ravens, namely “stand around near the roads so they can eat what is killed there.” The similarity between them and widows is “poor” (Shu,

2002, p.105). Therefore, understanding this metaphor correctly requires relevant knowledge background. ChatGPT failed to identify the similarity between raven and widow because it lacked the relevant knowledge background when processing the sentence.

• Induction of Metaphor Types

The test group used an excerpt from Shakespeare's Sonnet 73. The poem contains 5 primary metaphors: 1) LIGHT IS A SUBSTANCE; 2) EVENTS ARE ACTIONS; 3) LIFE IS A PRECIOUS POSSESSION; 4) A LIFETIME IS A DAY; 5) LIFE IS LIGHT (Shu, 2000, p.23).

In response to questions about the various conceptual metaphors in the selected excerpt, ChatGPT first answered the author and source of the poem and pointed out that the structure of the poem compares human life to the passing of time and the changing of seasons. ChatGPT then identified some of the conceptual metaphors in the poem, such as life as a day, aging as the fading of light, and death as a sleep. ChatGPT considered the first metaphor, life as a day, to be the foundation of the poem. Afterwards, we asked ChatGPT to provide more conceptual metaphors, and ChatGPT supplemented new metaphors such as the passing of time as a journey, the changing of seasons as aging, and second person address as addressing death.

Compared to the 5 primary metaphors mentioned earlier, we considered that ChatGPT's answers only corresponded to two of them, A LIFETIME IS A DAY and LIFE IS LIGHT. For example, "death as a sleep" can be categorized as A LIFETIME IS A DAY; "the passing of time as a journey" can be categorized as LIFE IS A JOURNEY; while "the changing of seasons as aging" essentially compares human life to a year and belongs to A LIFETIME IS A DAY. On the other hand, "second-person address as addressing death" essentially

metaphorizes a time point using spatial location. Therefore, the accuracy rate of this test group was about 40%.

• Metaphorical Inference

The corpus used in this test was divided into three groups as follows. The first group used a question from Ohori (2002, p.80), which compared the city to a body part, introduced a metaphor through a conditional sentence, and then explained the inference based on the metaphor in the following text. ChatGPT's answer completely matched the original text and provided a reasonable basis for the inference (i.e., "Just as the heart pumps blood throughout the body, the Seine River flows through the city, connecting its various neighborhoods and providing life to the city.").

The five single sentences in the second group were all examples of using the JOURNEY metaphor to describe LOVE. For each of the five sentences, ChatGPT appropriately understood the meaning of the sentence. In particular, the answers to (6), (7), and (9) were all based on the same conceptual metaphor. For example, in answering (6), ChatGPT's expression was "it may be helpful to take a step back and assess the situation"; in answering (7), ChatGPT's expression was "If... you're not sure how to get it back on course"; in answering (9), ChatGPT's expression was "If you have reached the point where you feel that ending the relationship is the best option". The italicized parts are all expressions of the journey metaphor.

The third group tested ChatGPT's reasoning competence in simulating scenarios. The text is from Lakoff (2007), and the metaphorical concept embodied in it is LOVE IS A JOURNEY. The inference provided by Lakoff (2007, p.153) is "Look, if it hit a dead-end-street, you can't keep going the way you have been going. You have to turn back." This advice is cleverly based on

the JOURNEY metaphor used by the female student to describe her problem. ChatGPT correctly identified the meaning of “dead-end-road”, that is, “their relationship has reached a point where there is no way to move forward or make progress”, and then offers the following advice: “work together to find a way to move past this dead end and move forward in their relationship.” Therefore, the accuracy rate of all three test groups was 100%.

• Discussion

In the conventional judgement tests, ChatGPT successfully passed five sets of pairwise comparison tests and provided reasonable justifications, demonstrating its competence to identify and understand both conventional and novel metaphors, and to evaluate the novelty of metaphors based on its understanding of them.

In the similarity understanding test, ChatGPT indicated that neither of the two given statements were semantically valid, exposing some issues in its competence to process input language. For instance, when processing statement (2), ChatGPT failed to activate relevant knowledge to identify possible similarities between the two objects, resulting in a deviation in its understanding of the metaphor.

In the induction of metaphor types test, we found that ChatGPT tended to extract the source and target domains of metaphors directly from the corpus, making it difficult for the model to perform induction at different levels of abstraction like humans do. Additionally, when extracting metaphor types from poetry, it tended to focus on the theme expressed by the poem, lacking sufficient flexibility.

In the metaphorical inference test, ChatGPT performed remarkably well, not only correctly understood the meanings of all the language materials, but

also provided reasonable answers for the first Group of inference materials (4), and demonstrated a certain degree of metaphorical reasoning competence by using the same conceptual metaphor as the test materials in the answers to the second Group of inference materials (6), (7), and (9).

• Output of Poem Generation

The poetry generated by ChatGPT used five conceptual metaphors, including LIFE IS A JOURNEY (see Appendix for details). For example, in the first section, life was compared to a “winding road” that people must navigate through “twists and turns, highs and lows,” which all represented the conventional metaphor of LIFE IS A JOURNEY. Similarly, the use of the word “balk” in the second section can also be seen as based on LIFE IS A JOURNEY. Therefore, it can be seen that ChatGPT’s output mainly relies on conventional metaphors when creating poetry on the theme of “man’s life.” As no novel metaphors were observed, the overall score for this output is two points.

• Output of Dialogue Generation

We provided the starting content of the dialogue and the corresponding conceptual metaphor, LOVE IS AN INVESTMENT, as shown in Table 8, and asked ChatGPT to write a continuation of the dialogue within six turns. ChatGPT ultimately produced a continuation consisting of four turns (see Appendix).

In the first turn of the continuation, the character Qin Fen expressed his understanding of LOVE IS AN INVESTMENT, namely that love should not only be based on the initial feeling when two people first meet, but also on future development. Stock Woman’s question was intended to ask about the advantages she has in attracting the opposite sex. Overall, the dialogue closely

followed the metaphor of LOVE IS AN INVESTMENT and used multiple mappings based on this metaphor, with clear and complete logical coherence. Stock Woman's use of "underlying value" to metaphorically described a person's inherent advantages, implying that this "underlying value" can offset the "risk" (i.e., shortcomings) of a person, was a relatively novel reasoning. Therefore, the score for the dialogue continuation was 3 points.

• Discussion

The metaphor generation competence test consists of two questions, with a maximum score of 3 points for each question, and a total score of 6 points. ChatGPT scored 5 points, indicating that it can generate a certain number of metaphorical expressions on its own. However, in terms of quantity, it tended to produce conventional metaphors and lacked novelty in metaphors. Therefore, generating novel metaphors may be a challenge for the current version of ChatGPT. Moreover, how to efficiently guide ChatGPT to generate more novel metaphorical expressions is a topic that needs to be further addressed in the future.

• Conclusion

Based on the previous research, this article has proposed a scheme for testing the metaphorical competence of intelligent chatbots and evaluated ChatGPT's performance in identifying, understanding, and generating metaphors. Overall, compared to previous artificial intelligence models, ChatGPT's metaphorical competence is indeed impressive (especially in metaphorical inference and generation), but there are still areas for improvement (such as low recognition rates for container metaphors and primary metaphors, as well as difficulties in

induction of metaphor types across different levels of abstraction).

In the future, it is worth exploring the potential application of cognitive linguistics and conceptual metaphor theory within the realm of language intelligence, further enhancing the language proficiency of artificial intelligence, and determining the comparative strengths and weaknesses of ChatGPT and humans in metaphor identification, comprehension, and generation.

References

- Bailey, R. (2003). Conceptual Metaphor, Language, Literature and Pedagogy. *Language Learning Journal*, 1(2), pp.59-72.
- Basso, K. (1976). *Wise words in western Apache*. Language, Culture and Cognition. Macmillan Publishing Co. Inc.
- Danesi, M. (1986). The role of metaphor in second language pedagogy. *Rassegna Italiana di Linguistica Applicata*, 18(3), pp.1-10.
- Danesi, M. (1993). Metaphorical competence in second language acquisition and second language teaching. In J. E. Alatis (Ed.), *Georgetown University Round Table on Language And Linguistics*. pp.489-500 Georgetown University Press.
- Danesi, M. (1994). Recent research on metaphor and teaching of Italian. *Italica*, 71(4), pp.453-464.
- Han, T., & Su, X. (2020). On the cognitive model of love metaphors in Japanese “enka”. *Advanced Japanese Language Education*(2), pp.80-89.
- Katz, A. N. (1996). On interpreting statements as metaphor or irony: Contextual heuristics and cognitive consequences. In J. S. Mio & A. J. Katz (Eds.), *Metaphor: Implications and Applications* (pp. 1-22). Psychology Press.
- Kovecses, Z. (2002). *Metaphor: A practical introduction*. Oxford University Press.
- Lakoff, G. (1993). The contemporary theory of metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202-251). Cambridge University Press.
- Lakoff, G. (2007). *Ten lectures on cognitive linguistics*. Foreign Language Teaching and Research Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.
- Lakoff, G., & Johnson, M. (2003). *Metaphors we live by*. University of Chicago Press.
- Li, Z., & Liang, G. (2022). Interdisciplinary construction of language intelligence. *Technology Enhanced Foreign Language Education*(5), pp.88-93, pp.117.

- Littlemore, J. (2001). Metaphoric competence: A language learning strength of students with a holistic cognitive style? *Tesol Quarterly*, 35(3), pp.459-491.
- Littlemore, J., & Low, G. (2006). Metaphoric competence, second language learning, and communicative language ability. *Applied linguistics*, 27(2), pp.268-294.
- Low, G. D. (1988). On teaching metaphor. *Applied linguistics*, 9(2), pp.125-147.
- Ohori, T. (2002). *Cognitive linguistics*. University of Tokyo Press.
- Pinker, S. (1994). *The language instinct*. The MIT Press.
- Shu, D. (1998). On the essence and semantic features of metaphor. *Journal of Foreign Languages*, (6), pp.10-19.
- Shu, D. (2000). Types of linguistic metaphors and their syntactic and semantic features. *Journal of Foreign Languages*, (1), pp. 20-28.
- Shu, D. (2002). On the working mechanism of metaphor. *Foreign Language Teaching and Research* (2), pp.98-106, pp.160.
- Wang, W. (2007). *Cognitive construction and interpretation over metaphors*. Shanghai Foreign Language Education Press.
- Zhong, Y. (2015). *The development and training of metaphoric competence for Japanese learners*. Wuhan University Press.
- Zhou, J., Lyu, X., Shi, J., & Kai, Z. [Research on Language Intelligence has gradually become a hot topic.] 『Chinese Social Sciences Today』 2017, February 7, (Viewed date: 5th. Apr. 2023)

Table 1. Pre-test

Items	Questions
Related Knowledge	<ol style="list-style-type: none"> 1. What is conceptual metaphor? How is it different from traditional rhetoric? 2. How does ChatGPT recognize metaphor? What are the specific steps?
Metaphor Identification	<ol style="list-style-type: none"> 1. Does the sentence “Our relationship has reached a dead end” contain conceptual metaphor? 2. Does the sentence “Behind his soft-spoken manner, the fire of ambition burned” contain conceptual metaphor?
Metaphor Comprehension	<ol style="list-style-type: none"> 1. Please explain the conceptual metaphor contained in the sentence “Behind his soft-spoken manner, the fire of ambition burned.” 2. What can this conceptual metaphor be summarized as?
Metaphor Generation	<ol style="list-style-type: none"> 1. Generate five sentences describing life that contain conceptual metaphors. 2. Generate a 100-word love poem that does not use the words “love” or “romance” and does not use figurative language.

Table 2. Test for Metaphor Identification Competence

Level	Group	Number	Description
Vocabulary	Words	5	Provide specific words and ask “Does this word have a metaphorical meaning or usage?”
	Phrases	5	Provide specific phrases (including idioms, etc.) and ask “Does this expression have a metaphorical meaning or usage?”
Sentence	Groups 1~6	27	<p>Provide single sentences for each group and ask “Does this expression contain a metaphorical expression or usage?”</p> <p>Group 1: Sentences based on conventional metaphors</p> <p>Group 2: Sentences based on container metaphors</p> <p>Group 3: Sentences with semantic conflicts</p> <p>Group 4: Sentences that violate metaphorical consistency</p> <p>Group 5: Sentences based on novel metaphors</p> <p>Group 6: Sentences based on primary metaphors</p>
Evaluation Criteria: Calculate the accuracy rate. If ChatGPT does not recognize the presence of a conceptual metaphor in the expression or judges improperly, it will be considered a failed recognition.			

Table 3. Test for Metaphor Comprehension Competence

Items	Number	Description
Conventional Judgement	10	Using 10 single sentences with novelty gradients, pairing two sentences from opposite ends of the gradient, and requiring judgment of the novelty of the two sentences within each pair.
Similarity Understanding	2	Given a single sentence, requiring judgment of whether the expression is reasonable or not.
Induction	1	Requiring the induction of metaphor types from the given corpus.
Inference	7	Conducting metaphorical inference based on the given corpus or answering questions in simulated real-life situations.
<p>Evaluation Criteria:</p> <p>① Conventional Judgment: Calculating accuracy based on existing research.</p> <p>② Similarity Understanding: Answering correctly if it matches existing research or providing a reasonable explanation.</p> <p>③ Induction: If it completely matches existing research, the accuracy is calculated as 100%; if it partially matches existing research, the accuracy is calculated proportionally; if it is completely unreasonable or the metaphorical expression is not recognized, the test is not passed.</p> <p>④ Inference: Passing the test if understood properly.</p>		

Table 4. Test for Metaphor Generation Competence

Items	Number	Description
Poetry	4	Provide a target domain and require the generation of at least 20 lines of conceptual metaphorical poetry around that target domain.
Dialogue	1	Provide the starting content of a dialogue and the corresponding conceptual metaphor, and require the continuation of a dialogue containing the specified conceptual metaphor.
Evaluation Criteria: According to Littlemore (2001), the stronger the ability to produce metaphors, the higher the score if the output is reasonable or more novel. For each output, if it contains one or more novel metaphors, a full score of 3 points is given. If completely unreasonable metaphors are used, the lowest score of 0 points is given.		

Table 5. Corpus for Metaphor Identification Competence at the Lexical-Level

Level		Test Corpus
Lexicon	Word	Crane; Father; Mouse; Sweetheart; Motherland
	Phrase	invisible hand; free lunch; safety net; glass ceiling; the foot of a hill

Table 6. Corpus for Metaphor Identification Competence at the Sentence-Level

	Test Corpus	Criteria
G1	<p>(1a) Is that the <i>foundation</i> for your theory?</p> <p>(1b) The theory needs more <i>support</i>.</p> <p>(1c) We need to <i>construct</i> a strong argument for that.</p> <p>(1d) The theory will <i>stand</i> or <i>fall</i> on the <i>strength</i> of that argument.</p> <p>(1e) So far we have <i>put together</i> only the <i>framework</i> of the theory. (Kövecses 2002, p.5)</p>	<p>The underlying metaphor in this group is THEORIES ARE BUILDINGS. The italicized words contain the metaphoric meanings. If ChatGPT judges a sentence as a metaphorical expression or successfully recognizes the key words, it is considered as passing the test.</p>
G2	<p>(2a) Are you <i>in</i> the race on Sunday?</p> <p>(2b) Are you <i>going to</i> the race?</p> <p>(2c) Did you <i>see</i> the race?</p> <p>(2d) The <i>finish</i> of the race was really exciting.</p> <p>(2e) There was a <i>lot of good running in</i> the race.</p> <p>(2f) I couldn't do <i>much sprinting</i> until the end.</p> <p>(2g) <i>Halfway into</i> the race, I ran out of <i>energy</i>. (Lakoff & Johnson, 2003, p.31)</p>	<p>The underlying metaphor in this group is the container metaphor. The italicized words contain the metaphorical meanings. If ChatGPT judges a sentence as a metaphorical expression or the basis for judgment is reasonable, it is considered as passing the test.</p>
G3	<p>(3) The old rock is becoming brittle with age. (Shu, 1998, p.13)</p>	<p>This sentence contains a semantic conflict. If ChatGPT judges this sentence as a metaphorical expression, it is considered as passing the test.</p>

G4	<p>(4a) *This building looks like a vast ocean.</p> <p>(4b) *Her long hair is like a pill.</p> <p>(4c) *The car sped along the winding mountain road like an arrow shot from a bow.</p> <p>(4d) *His ears look like long long roads, along which trucks can run.</p> <p>(4e) *The book is a bright moon that hangs over the stormy sea (Wang, 2007, p.100).</p>	<p>All sentences in this group violate the principle of metaphorical consistency and cannot stand. If ChatGPT judges that a sentence is not valid and the explanation is reasonable, it is considered as passing the test.</p>
G5	<p>(5a) His theory has thousands of little rooms and long, winding corridors.</p> <p>(5b) His theories are Bauhaus in their pseudo functional simplicity.</p> <p>(5c) He prefers massive Gothic theories covered with gargoyles.</p> <p>(5d) Complex theories usually have problems with the plumbing (Lakoff & Johnson, 2003, p.53).</p>	<p>These sentences belong to the imaginative expression of THEORIES ARE BUILDINGS. If ChatGPT judges that a sentence contains metaphorical expressions or provides a reasonable basis, it is considered as passing the test.。</p>
G6	<p>(6a) They greeted me <i>warmly</i>. (Affection Is Warmth)</p> <p>(6b) Tomorrow is a <i>big day</i>. (Important Is Big)</p> <p>(6c) I'm feeling <i>up</i> today. (Happy Is Up)</p> <p>(6d) This movie <i>stinks</i>. (Bad Is Stinky)</p> <p>(6e) Prices are <i>high</i>. (More Is Up) (Lakoff & Johnson, 1999, pp.50-51)</p>	<p>This group of sentences tests the ability to recognize primary metaphors. The italicized words in the sentences contain the key words indicating the presence of metaphorical meanings.</p>

Table 7. Corpus for Metaphor Comprehension Competence Test

Items	Group	Test Corpus
Conventional Judgment	A	(A1) Handshakes are doorbells. (A2) Children are precious gems (Katz, 1996, p.13).
	B	(B1) Their love is a coniferous tree. (B2) That experience was a crushing blow. (ibid:13)
	C	(C1) Their friends are houseplants. (C2) She is a magnet to men. (ibid:13)
	D	(D1) That comment has the bull's eye. (D2) This textbook is a winning lottery ticket. (ibid:13)
	E	(E1) His standards are at the ceiling. (E2) His smiles are can openers. (ibid:13)
Similarity Understanding		(1) Metabolically, the brain is a pig (Pinker, 1994). (2) Ravens are widows (Basso, 1976, p.249).
Induction of metaphor types		(3) In me thou seest the twilight of such day As after sunset fadeth in the west; Which by and by black night doth take away, Death's second self that seals up all in rest. (Shakespeare Sonnet 73)

Metaphorical Inference	G1	(4)If Notre Dame is the heart of Paris, then what is its aorta ? (Ohori, 2002, p.80)
	G2	(5) We're at a crossroads. (6) We're spinning our wheels. (7) Our relationship is off the track. (8) The marriage is on the rocks. (9) We may have to bail out of this relationship. (Lakoff, 1993, p.206)
	G3	(10) "I'm sorry. I have a metaphor problem with my boyfriend." And she said, "Maybe, you can help." So she then said, "On the way here, my boyfriend said that our relationship had hit a dead-end-street." And she didn't know what this meant. (Lakoff, 2007, p.153)

Table 8. Corpus for Metaphor Generation Competence Test

Items	Test Corpus
Poetry	Generate a poem to describe a man's life in 20 lines.
Dialogue	<p>Qin Fen is on a blind date with a stock woman. Following is their dialogue.</p> <p>Qin Fen: Then I should be considered a stock with what performance?</p> <p>The stock woman: from the age and looks, it should belong to the kind of falling below the issue price, right?</p> <p>Continue this dialogue in 6 turns from the perspective of conceptual metaphor "love is an investment". (Han & Shu, 2020, p.86)</p>

Table 9. Evaluation Criteria for Metaphor Generation Competence

Score	Criteria
3: Novel Metaphor	Meets at least one of the following conditions: (a) The topic is projected onto a sensory domain that is generally not applicable, resulting in a metaphor that is not a familiar English expression. (b) The topic is often associated with the physical world but is projected onto a mental state (or vice versa), resulting in a metaphor that is not a familiar English expression. (c) The topic remains within its habitual domain, but there is a significant change in perspective, resulting in a metaphor that is not a familiar English expression.
2: Conventional Metaphor	The produced metaphor is a familiar English expression.
1: Literal Meaning	The topic remains within its habitual domain.
0: Inappropriate Expression	Both evaluators are unable to find any meaningful metaphorical expression.

Appendix: Results of Metaphor Generation Competence Test

Items	Dialogue (Bold indicates the test-taker, others are ChatGPT)
Poetry Generation	<p>A poem to describe a man’s life in 20 lines please.</p> <p>A man’s life is like a winding road, /With twists and turns that he must goad, /Through highs and lows, he makes his way, /And faces challenges every day. /He learns to crawl and then to walk, /He learns to talk and then to balk, /He meets new friends and falls in love, /And looks to the stars above. /He grows up fast, with dreams to chase, /And leaves his childhood in its place, / He works hard, with sweat and tears, /And overcomes his doubts and fears. /He finds success and feels alive, /And takes each moment to thrive, /He shares his joy and gives his heart, /And makes a difference with his part. /He ages gracefully, with grace and poise, /And treasures memories of his joys, /He looks back on his life’s design, /And knows he lived it well, divine.</p>

<p>Dialogue Generation</p>	<p>Qin Fen is on a blind date with a stock woman. Following is their dialogue.</p> <p>Qin Fen: Then I should be considered a stock with what performance?</p> <p>The stock woman: from the age and looks, it should belong to the kind of falling below the issue price, right?</p> <p>Continue this dialogue in 6 turns from the perspective of conceptual metaphor “love is investment”.</p> <p>Qin Fen: I see. Well, I believe that love is like an investment. It’s not just about the initial appearance, but also about the underlying value.</p> <p>Stock woman: That’s an interesting way to think about it. So, how would you rate your own value as an investment in love?</p> <p>Qin Fen: I like to think that I have a lot of potential for growth and development in a relationship. Just like a stock with a solid foundation, I have a strong character and a good heart.</p> <p>Stock woman: That’s reassuring to hear. But just like with any investment, there are always risks involved. What are some potential risks in investing in a relationship with you?</p> <p>Qin Fen: Well, I suppose one risk could be that I’m not always the most communicative person. Sometimes I need a little push to open up and share my feelings.</p> <p>Stock woman: I see. Communication is definitely an important factor in any relationship. But if the underlying value is strong, I believe that there is always room for growth and improvement.</p> <p>Qin Fen: Absolutely. And just like with stocks, it’s important to do your research and make informed decisions. I’m always willing to put in the effort to make a relationship work and thrive.</p> <p>Stock woman: I appreciate your dedication to investing in a relationship. With the right balance of risk and reward, I believe that we could have a profitable future together.</p>
--------------------------------	---

