ISSN 1798-4769
Journal of Language Teaching and Research, Vol. 16, No. 5, pp. 1665-1672, September 2025
DOI: https://doi.org/10.17507/jltr.1605.24

Can ChatGPT Analyze Textual Data? The Case of Conceptual Metaphors in Short Stories of Language Assessment*

Hui Geng[1]
College of Foreign Studies, Guilin University of Technology, Guilin, China

Han Wei
College of Foreign Studies, Guilin Institute of Information Technology, Guilin, China;
School of Education, Taylor's University Lakeside Campus, Subang Jaya, Malaysia

Vahid Nimehchisalem
Faculty of Humanities and Arts, University of Economics and Human Sciences in Warsaw, Warsaw, Poland;
School of Education, Taylor's University Lakeside Campus, Subang Jaya, Malaysia

Ali Sorayyaei Azar
Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia

***Abstract*—ChatGPT, a modern artificial intelligence (AI) chatbot, has emerged as an unprecedented breakthrough in multiple domains traditionally dominated by humans. Its ability to engage in human-like conversations has the potential to influence the fields of linguistics and education. The fundamental functions of ChatGPT in teaching and learning have been the subject of some research, but its application in textual analysis has received scant attention. This study aims to investigate how ChatGPT assists in analyzing conceptual metaphors (CMs) in short stories used in language assessment. Based on the Conceptual Metaphor Theory (CMT) by Lakoff and Johnson (1980), the study identified the structural, orientational, and ontological metaphors in 22 short stories from the book *Tests and Us 2*, first by the cutting-edge AI program ChatGPT (GPT-4), then refined by the researchers and validated by linguistic experts. The results showed a total of 250 conceptual metaphors, including 131 structural metaphors, 64 ontological metaphors, and 55 orientational metaphors. When validated by human specialists, GPT-4 accurately recognized conceptual metaphors in 81.2% of the cases, amounting to 203 instances. The most dominant error made by GPT-4 was classifying non-metaphoric expressions as metaphoric, followed by providing unclear explanations and classifying metaphoric expressions as non-metaphoric. Errors associated with being too general and too non-literal, unmatched categorization as well as wrong mapping of source or target domain also occurred. Our study shows that ChatGPT, despite its controversial position in academic settings, can be used as a relatively reliable tool in aiding the analysis of textual data.**

***Index Terms*—ChatGPT, conceptual metaphors, textual analysis, short stories, language assessment**

I. Introduction

ChatGPT, an artificial intelligence chatbot that employs natural language processing to simulate human conversation, has recently sparked widespread interest. It is capable of responding to inquiries and producing a wide range of written materials, like articles, essays, emails, poetry, social media postings, and so on, using the prompts offered on the interface at the request of its human interlocutors. According to Grimmer and Stewart (2013), automated techniques have long been employed by researchers to facilitate the analysis of textual data, thereby enabling the extension of content analysis to corpora of larger sizes that were previously unattainable through manual means. The emergence of Chatbot has further expanded the application of automated content analysis. As one of the Generative Large Language Models (GLLMs), ChatGPT also has the potential to conduct textual analysis in a variety of fields to assist humans. Previous research demonstrated its effectiveness in accounting and compared it to other existing textual analysis approaches (de Kok, 2024). However, there is a scarcity of studies on text analysis with the use of ChatGPT in the field of applied linguistics, particularly in analyzing conceptual metaphors. Therefore, this paper aims to explore the possibility of employing ChatGPT as an instructional and coding tool in analyzing conceptual metaphors, with the primary objective of investigating its capability to analyze conceptual metaphors in short stories of language assessment.

* This paper is the outcome of the research project RD2500000526.
[1] Corresponding Author.

A case study was conducted by establishing the corpus with 22 short stories of language assessment. Following the Conceptual Metaphor Theory (CMT) by Lakoff and Johnson (1980), the structural, orientational, and ontological metaphors were identified and analyzed by ChatGPT (GPT-4 version). Though textual analysis through chatbot will never replace the careful and close reading of texts by humans, the results of conceptual metaphors identified and analyzed by ChatGPT have yielded some insights in supporting manual analysis. Future studies will benefit from its powerful, adaptable, fast, and cost-effective nature.

II. Literature Review

The Generative Pre-trained Transformer (GPT), commonly referred to as ChatGPT, is a type of Large Language Model (LLM) and a prominent framework for generative artificial intelligence (AI). Since 2018, OpenAI has unveiled several generations of the GPT as a large language model, followed by the GPT-2 and GPT-3 generations in 2019 and 2020 respectively. Depending on the development of GPT, the parameters available to deal with would change. Notably, the GPT-3 model is distinguished by its impressive 175 billion parameters (Zhu et al., 2023), rendering it the most extensive language model to have undergone training. In terms of conversational AI chatbots, the GPT-3 was launched as the ChatGPT, which could respond in a coherent and contextually correct way (Qammar et al., 2023). GPT technology has evolved, and by the time ChatGPT became accessible to the public in November 2022, it reached version 3.5, attracting significant public and academic interest (Reiss, 2023). In tests of human-machine interaction, it has demonstrated remarkable natural language comprehension skills and exceptional proficiency in managing substantial volumes of textual data (Tan et al., 2023). GPT-4, unveiled by OpenAI on March 14, 2023, is the recent version of OpenAI's GPT series. It is an enhanced version of GPT-3.5 with the superior capacity to identify nuances and deliver more accurate and logical responses (Cohen & Adams, 2024).

ChatGPT provides the prospect of replacing labor-intensive and time-consuming work by either annotating text or classifying texts (Reiss, 2023). Tasks that would previously take months might take only a couple of weeks to solve (de Kok, 2024). Among many capabilities, GPTs can be employed for linguistic tasks to generate promising results, such as in machine translation, text generation and summarization, dependency parsing, and textual similarity analysis (ÜNLÜ, 2023). Researchers in the field of linguistics have also tried creating their own algorithms for analyzing data. For instance, Hassan et al. (2019) used a large training dataset of political texts to train an algorithm to identify feedback about political risk in their quarterly earnings conference calls. Coders' dream in applied linguistics research has always been accomplishing such cumbersome and time-consuming tasks in more feasible ways. ChatGPT may have some potential for textual analysis. However, despite the value of ChatGPT in textual analysis, there has been a limited number of researchers who have undertaken case studies in the domain of applied linguistics. Our objective in this study is to examine whether ChatGPT can recognize metaphors in certain genres and texts, especially in creative writing.

Metaphors are extensively prevalent in all facets of our language, to the extent that achieving effective communication becomes nearly unattainable in their absence. From a literary standpoint, it can be argued that metaphors serve the purpose of enhancing the vividness and aesthetic appeal of language. According to Lakoff and Johnson (2008), the essence of metaphor is understanding one thing in terms of another thing. In their book entitled *Metaphors We Live By*, they mentioned that metaphor was originally not a linguistic process but a process of mind, and they labeled it the conceptual metaphor. The conceptual metaphors are derived from personal experience and perception of the surrounding world. Lakoff and Johnson's (1980) seminal work on metaphors provided the theoretical foundation for subsequent work on metaphors in cognitive linguistics, discourse analysis, and communication studies (e.g., Gibbs, 2011; McGlone, 2007). Early research by Lakoff and Johnson (2008) led to the development of a set of theories known as Conceptual Metaphor Theory (CMT), in which structural, orientational, and ontological metaphors were presented and distinguished.

According to Lakoff and Johnson (1980), in structural metaphor, complicated and abstract experiences are conceptualized based on the experience of simple and specific experiences. An example often used to illustrate this type is "ARGUMENT IS WAR". In orientational metaphor, a system of ideas is organized in the relation and interaction in space like up-down, inside out, front-behind, shallow-deep, center-periphery, etc. An example often cited to illustrate this type is "HAPPY IS UP, SAD IS DOWN". In ontological metaphor, an abstraction, such as an activity, emotion, or idea, is represented as something concrete, such as an object, substance, container, or person. A typical example is "THE MIND IS A MACHINE". These three types of conceptual metaphors often occur in our daily life communications as well as story writing.

Since stories can be interpreted in several ways and need imaginative re-creations of the real world, they are not just a chronological account of what happened (Kearney, 2002). According to Boje (2017), stories are viewed as the preferred sense-making currency of human relationships among internal and external stakeholders. Stories have the capacity to convey real processes in a literal and factual manner, while also utilizing conceptual metaphors to elicit vivid imagery and encapsulate the ongoing flow of events associated with an organization (Gray, 2007). In the teaching and learning setting, instructor effectiveness and student engagement can be improved through storytelling (Baruch, 2006; Hibbert, 2013; Schultz & Quinn, 2014; Taylor & Statler, 2014).

The ideas and themes given by the stories will be better appreciated if the structural, orientational, and ontological metaphors are noted. While there have been prior investigations into conceptual metaphors in a variety of stories

(Bednarek, 2005; Anh, 2017; Taylor & Statler, 2014), the utilization of AI technology for aiding textual data analysis has not yet been integrated. Prior research has employed traditional methods of metaphor identification, including intuition, dictionary-based approaches, the Metaphor Identification Procedure (MIP or MIPVU) as proposed by Group (2007), and the Procedure for Identifying Metaphorical Scenes (PIMS) developed by Johansson Falck and Okonski (2023). However, no study has directly analyzed metaphors using GPT-4. Meanwhile, due to the inherent human nature of researchers, previous studies have employed the usage of corpora and selectively examined specific typical metaphors to serve their own descriptive purposes concerning certain themes, rather than undertaking a comprehensive analysis of all conceptual metaphors and then extracting typical metaphors in a book. Therefore, the current study aims to fill this gap by using ChatGPT (GPT-4) to identify the three types of conceptual metaphors in a self-build corpus with 22 short stories based on the theme of language assessment.

III. Methods

A. Corpus Establishment

The sample of this study comprised 22 short stories sourced from a published book titled *Tests and Us – A Collection of Real Stories (Volume 2)* (Nimehchisalem & Geng, 2023). The textual data were collected by self-establishing a corpus consisting of 22 short stories related to language assessment. The total word count amounted to approximately 23,000 words, with an average of 1,045 words per story.

B. Instruments

The present study employed an AI-assisted approach for qualitative analysis of textual data. The AI-assisted approach aimed to examine conceptual metaphors present in narrative pieces related to language assessment, utilizing the popular AI program ChatGPT (GPT-4). In the meantime, the outcomes of conceptual metaphors were evaluated via manual analysis conducted by human coders using traditional coding techniques.

C. Data Analysis

The first step in the data analysis process was machine coding. To investigate the capacity of ChatGPT for analyzing conceptual metaphors within these stories, the GPT-4 version was installed on the Windows operating system. Subsequently, after clicking the "new chat" button, the researchers successfully engaged in dialogue with ChatGPT by asking questions and conversationally providing prompts. When the dialogue interface was blank, the researchers could enter a story. Then, conversations began.

The first question posed to ChatGPT was:

- *Story XXX*
- *Could you please analyze all the conceptual metaphors (structural, orientational, and ontological metaphors) in the short story and categorize them into a table with columns of conceptual metaphors, categories, examples, and explanations?*

Shortly it answered:

- *Certainly! Below is the analysis of the conceptual metaphors in the provided text, categorized into a table:…*

Here, the results and a table with the required columns were attached. If the researchers deemed the outcomes to be satisfactory, we would proceed with the action of copying and pasting the table from the dialogue interface into a Word document. If the researchers perceived the outcomes to lack comprehensiveness, we would choose to either reiterate the same previous inquiry or emphasize the point by inserting the question into ChatGPT as:

- *Could you please comprehensively analyze all the conceptual metaphors (structural, orientational, and ontological metaphors) in this short story and categorize them into a table with columns of conceptual metaphors, categories, examples, and explanations?*

Then, ChatGPT would answer:

- *Certainly. Below is a more comprehensive analysis of the conceptual metaphors from the given text:…*

Sometimes at the end of the conversation, ChatGPT would add a disclaimer or a summary of the analysis:

- *Please note that the explanations provided here are based on common interpretations and conceptual mappings of the metaphors in the text. Different readers may have varying interpretations based on their perspectives and experiences.*
- *These metaphors provide a rich texture to the story, adding depth to the characters and creating a more vivid and relatable narrative.*

To our observation, the results generated by ChatGPT for the second time would meet the researchers' standards. If the researchers feel that there should be more metaphors, then we could ask ChatGPT:

- *Any more structural, orientational, and ontological metaphors in this short story?*

And ChatGPT would patiently answer again:

- *Certainly, I can further identify some more subtle conceptual metaphors in the short story.*

If the researchers had any doubts, then we could keep discussing and collecting results.

Then, the next question would follow. The second question posed to ChatGPT was:

- *Could you please count the frequency of these three types of metaphors and put them into a table?*

Shortly it answered:

- *Certainly. Here's the count for each category of conceptual metaphors from the analysis: (Table).*

Based on the results, it showed a brief description as well:

- *The table shows that structural metaphors are the most common in the text, followed by ontological metaphors, and finally orientational metaphors.*

In this way, the data analyzed by ChatGPT was ready within minutes.

The following step in the data analysis process was human coding. The researchers invited an expert in the area to evaluate the results that had been analyzed and reported by ChatGPT. This human coder was a linguist, who was experienced in the area and had analyzed conceptual metaphors and published journal articles related to conceptual metaphors. She checked the conceptual metaphors generated by GPT-4, and then the researchers checked the type of errors made by GPT-4 and made a discussion.

IV. Results and Discussion

A. Capability and Limitations of ChatGPT in Analysing Conceptual Metaphors

Based on the results of the human coder's examination of the results generated by ChatGPT, we realized that ChatGPT is capable of analyzing conceptual metaphors following the researchers' request, with the process consisting of six major phases: preprocessing of input texts – utilization of Language Model Architecture – understanding of language structure and semantics – recognition of conceptual metaphors – comprehension of context – inference and response.

During the initial stage, ChatGPT operates preprocessing procedures to transform the raw input texts into a format that is appropriate for its language model. The process commonly entails tokenization, wherein the 22 short stories are partitioned into smaller meaningful units for the model to understand. These smaller units are referred to as tokens (such as words, subwords, or characters), and numerical values are allocated to each token.

In the second phase, ChatGPT employs a transformer-based language model (Language Model Architecture), which is a neural network specifically designed for natural language processing tasks, and enables it to capture contextual dependencies and relationships between tokens in the text. As mentioned by Choi and Lee (2023), transformer models can capture the dependencies between inputs and outputs without regard for their distance in the sequence. Hence, Language Model Architecture facilitates a deeper understanding of the content of stories. During the third stage, ChatGPT undergoes thorough instruction using a varied dataset that encompasses a wide range of text genres, including narratives such as the specific example presented.

During the training process, the model acquires the ability to estimate the probability of a token's occurrence based on the contextual information provided by other tokens inside a sentence or document. This procedure facilitates the model's acquisition of comprehension regarding linguistic patterns and semantics, hence enhancing its ability to identify metaphors and comprehend textual information with greater efficacy.

The fourth phase occurs when ChatGPT uses its learned knowledge of language patterns to recognize metaphorical expressions. Conceptual metaphors involve mapping one concept (source domain) onto another concept (target domain) to convey meaning. The model identifies such mappings by analyzing the linguistic patterns and their context. For example, during the analysis of the short story *Unforgiven* (Nimehchisalem, 2023) which is related to unjust punishment, the metaphorical expressions "stone face" and "killer whale attack" are recognized as figurative representations that convey abstract ideas using concrete terms.

Next, the fifth phase comes when ChatGPT takes into account the words and phrases preceding and following the metaphorical expression to grasp its intended meaning. For instance, the context surrounding the "stone face" metaphor indicates the stern and intimidating demeanor of the teacher, Mr. Ghasemi. The model uses attention mechanisms in the transformer architecture to weigh the importance of different words in the context.

Finally, based on the analysis, ChatGPT formulates its response, providing insights and observations regarding the story's content and the use of conceptual metaphors. It acknowledges the potential reasons behind Mr. Ghasemi's behavior while leaving room for different interpretations and perspectives on the matter.

In addition to the capabilities of ChatGPT in analyzing conceptual metaphors in short stories, there are also some limitations and challenges.

The first limitation lies in the coding process. Unlike human coders, ChatGPT codes conceptual metaphors based on the basic and common Conceptual Metaphor Theory (CMT). The updated or adapted analytical framework by the most recent studies for conceptual metaphor analysis might not be fully processed and handled by ChatGPT. For example, the phrase "the beautiful music whistling in my ears" used in the short story *A Seed in a Desert or Oasis* (Geng, 2023) can be seen as an example of a synesthetic metaphor that mixes different sensory experiences, leading to a more vivid description. Here, the act of producing sharp and clear sounds in the source domain "whistling" is mapped onto the quality of the "music" in the target domain. This helps to convey a sense of the music's nature or its effect on the listener, making it more vivid and tangible. Therefore, this metaphor is beyond the basic conceptual metaphor framework of structural, orientational, or ontological categorization.

Second, ChatGPT encodes structural, orientational, and ontological metaphors based on general awareness. Consequently, there may exist divergent interpretations contingent upon the unique viewpoints and experiences of

individual readers. In the case of "classroom is a desert or oasis" from the short story *A Seed in a Desert or Oasis* (Geng, 2023), GPT-4 categorized it as a structural metaphor. However, the researchers believe that the metaphor used is more of an ontological metaphor than a structural one, as it describes the abstract concept of a classroom experience in terms of a physical, concrete entity, such as a desert or an oasis. The writer sharply contrasts two vastly different educational settings. Just as a desert is barren, hostile, and stifling, symbolizing a classroom lacking resources, feeling sterile, and suppressing creativity, an oasis represents the opposite. An oasis, being nurturing, revitalizing, and abundant, mirrors a classroom that promotes intellectual growth, offers relief from the pressures of learning, and fosters a flow of ideas. This metaphor highlights the significant influence that classroom environment and teaching methods can have on a student's eagerness to learn.

Third, due to the transformer-based language model's lack of explicit knowledge about the cultural context and specific historical references, it is hard for ChatGPT to interpret culturally specific metaphors and make explanations on historical allusions. For example, in the short story *Unforgiven* (Nimehchisalem, 2023), one of the challenges is comprehending the cultural implications of viewing "war as a hostile environment" and being a "war child". This metaphor alludes to the difficult circumstances of growing up during a time of war, where the psychological impact on the children and their learning experiences might be affected. ChatGPT might find it challenging to fully grasp the emotional weight and complexities associated with such an experience without a comprehensive understanding of historical contexts. Moreover, the mention of "Julius Caesar and his nephew" pertains to historical references that ChatGPT may not possess. The metaphor involves the reordering of months in the calendar, which is a reference to the Julian calendar reforms by Julius Caesar and the later Julian-Gregorian calendar changes. Without historical knowledge, ChatGPT might not be able to fully comprehend the significance and impact of these references.

B. Feedback From the Human Coder About the Accuracy of the Coding Generated by ChatGPT

Figure 1 illustrates the findings from ChatGPT (GPT-4) regarding conceptual metaphor analysis. Out of 250 conceptual metaphors detected, structural metaphors took the lead with 131 instances, accounting for 52.4%. This was trailed by ontological metaphors, which had 64 instances or 25.6%, and orientational metaphors, which made up 22.0% with 55 instances. These metaphors illuminated the underlying structure of the mind by providing the link between the body and the outside environment. Through conceptual metaphors, readers can grasp the complex layers of meaning that writers embed, leading them to a cognitive exploration of the sub-themes of language assessment stories in the future.

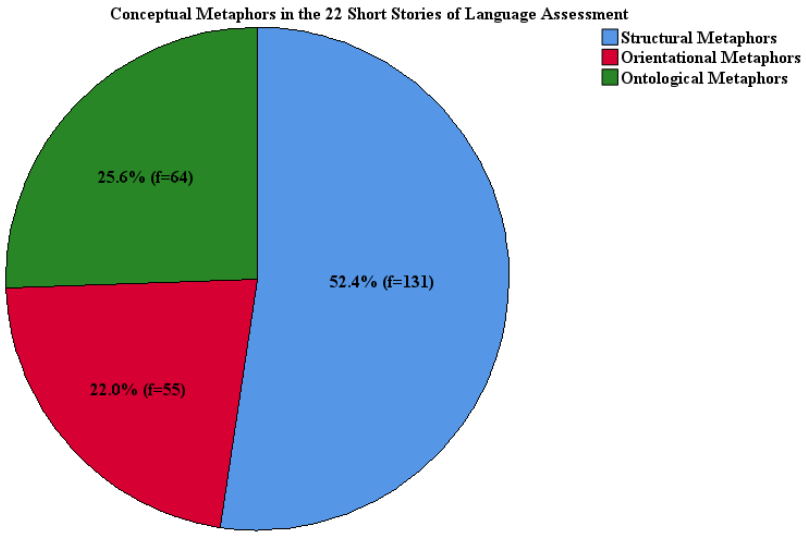

Figure 1. GPT-4 Results of Conceptual Metaphors in 22 Short Stories of Language Assessment

In previous research, human coders manually examined the results generated by GPT-3 variants for metaphors in Lakoff's Master Metaphor List and classified the errors as follows: wrong with trigger, wrong without trigger, too literal, should be non-metaphoric, should be metaphoric, too specific, too general, wrong subelement mapping. Trigger in this context refers to terms in the input that are unambiguously associated with the predicted source domain.

As a more advanced GPT-4 program was used to analyze metaphors in the current study, human coders verified the accuracy of the results. Table 1 presents the validation results, including the frequency and percentage of accurate conceptual metaphors identified by GPT-4, as well as the types of errors it made during the analysis of 22 short stories of language assessment.

TABLE 1
VALIDATION FOR RESULTS OF CONCEPTUAL METAPHORS ANALYSIS BY GPT-4

Types of Errors by GPT-4	Frequency	Percentage
1. Unmatched categorization	3	1.2%
2. Should be non-metaphoric	21	8.4%
3. Should be metaphoric	5	2.0%
4. Too general	4	1.6%
5. Too specific	0	0%
6. Too literal	0	0%
7. Too non-literal	4	1.6%
8. Wrong mapping of Source or Target Domain	1	0.4%
9. Unclear explanations	9	3.6%
Total Errors by GPT-4	47	18.8%
Clear Conceptual Metaphors by GPT-4	203	81.2%

The table categorizes the errors made by GPT-4 into nine distinct types. Of all the conceptual metaphors analyzed, GPT-4 made errors in 18.8% of the cases or 47 instances. The most prominent error was classifying non-metaphoric expressions as metaphoric, with 21 occurrences amounting to 8.4% of all evaluations. This indicates the potential over-sensitivity of the model to metaphoric interpretations. This might be a result of GPT-4's extensive training data, which could make it more prone to seeing patterns even where they might not exist. Unclear explanations were the next common type, with 9 instances at 3.6%. Meanwhile, GPT-4 sometimes classified metaphoric expression as non-metaphoric, with 5 instances at 2.0%. Other noticeable error types like being too general, too literal, and unmatched categorization also occurred.

On the brighter side, GPT-4 identified clear conceptual metaphors with an accuracy rate of 81.2%, or 203 instances. The results from Table 1 suggest that GPT-4 displays a commendable proficiency in analyzing conceptual metaphors, with over 80% accuracy. Such a high accuracy rate, especially in a nuanced task such as metaphor analysis, showcases the capabilities of GPT-4. These results signified the complexity of metaphor analysis and highlighted the importance of human validation in such tasks, particularly in academic and professional settings.

V. CONCLUSION

This study spotlighted the untapped potential of ChatGPT in facilitating the analysis of conceptual metaphors in short stories of language assessment. Based on the findings, ChatGPT is capable of analyzing textual data by engaging in and providing human-like conversations. Taking Lakoff and Johnson's (1980) Conceptual Metaphor Theory (CMT) as a basic analytical framework, a total of 250 conceptual metaphors are detected by GPT-4, including 131 structural metaphors, 64 ontological metaphors, and 55 orientational metaphors. Based on the validation of results by human experts, GPT-4 identified clear conceptual metaphors with an accuracy rate of 81.2%, or 203 instances. These metaphorical constructs offer illuminating insights into the intricate interplay between cognitive structures and external environments, thereby providing a deeper understanding of the cognitive underpinnings of the human mind. Moreover, they offer readers a lens to relate to abstract concepts, drawing them into an emotionally resonant narrative. Through metaphors, writers can craft vivid imagery, set the tone, and embed societal commentary, all while showcasing their unique stylistic voice. In essence, these metaphors enrich a short story, providing depth and layered meanings that resonate differently with each reader.

However, the challenges that ChatGPT encountered lie in the coding and identification process. It has difficulty coding conceptual metaphors on the updated or adapted analytical framework, and sometimes it fails to distinguish cultural and historical elements associated with metaphors. In the current study, several types of errors were found in making conceptual metaphor analyses by GPT-4. The most dominant error made by GPT-4 was classifying non-metaphoric expressions as metaphoric, followed by the provision of unclear explanations and the classification of metaphoric expressions as non-metaphoric. Errors associated with being too general and too non-literal, unmatched categorization as well as wrong mapping of source or target domain also occurred. Given these findings, it is recommended that future iterations of ChatGPT or similar models undergo more refined training processes that focus on the accurate detection of metaphoric content. Special attention should be given to prevent the misclassification of non-metaphoric expressions. Introducing an updated database with a comprehensive range of cultural and historical contexts can improve the model's performance in recognizing the subtleties within metaphors. Moreover, collaborating with linguists and cultural experts during the training phase might enhance the model's sensitivity to intricate details and nuances.

The relatively small size of the corpus was a constraint of this study. More stories and texts could be added in the future. Meanwhile, it is feasible to analyze conceptual metaphors in other literary genres such as poetry, prose, and drama with the emerging AI tools. Comparisons of the coding and identification of conceptual metaphors can also be made between ChatGPT and human experts.

REFERENCES

[1] Anh, D. T. M. (2017). An investigation into conceptual metaphors denoting life in American and Vietnamese short stories. *Journal of Development Research*, *1*(1), 29-35. https://doi.org/10.28926/jdr.v1i1.16

[2] Baruch, Y. (2006). Role-play teaching acting in the classroom. *Management Learning*, *37*(1), 43–61. https://doi.org/10.1177/1350507606060980

[3] Bednarek, M. A. (2005). Construing the world: Conceptual metaphors and event construals in news stories. *Metaphorik.de.* (pp. 6-32).

[4] Boje, D. M. (2017). The storytelling organization: A study of story performance in an office-supply firm. In S. Minahan (Ed.), *The aesthetic turn in management* (pp. 211-231). Routledge.

[5] Choi, S. R., & Lee, M. (2023). Transformer architecture and attention mechanisms in genome data analysis: A comprehensive review. *Biology*, *12*(7), 1033. https://doi.org/10.3390/biology12071033

[6] Cohen, J. D., & Adams, D. (2024). Text understanding in GPT-4 versus humans. *Royal Society Open Science*, *11*(5), 241313. https://doi.org/10.1098/rsos.241313

[7] de Kok, T. (2024). ChatGPT for textual analysis? How to use generative LLMs in accounting research. *Management Science, Forthcoming.* Available at http://dx.doi.org/10.2139/ssrn.4429658

[8] Geng, H. (2023). *A Seed in a Desert or Oasis*. In Nimehchisalem, V., & Geng, H. (Eds.). *Tests & Us - A Collection of Real Stories* (Vol. 2). (pp. 15-19). Generis Publishing.

[9] Gibbs Jr, R. W. (2011). Evaluating conceptual metaphor theory. *Discourse Processes*, *48*(8), 529-562. https://doi.org/10.1080/0163853X.2011.606103

[10] Gray, D. E. (2007). Facilitating management learning developing critical reflection through reflective tools. *Management Learning*, *38*(5), 495–517. https://doi.org/10.1177/1350507607083204

[11] Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267-297. https://doi.org/10.1093/pan/mps028

[12] Group, P. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, *22*(1), 1-39. https://www.tandfonline.com/doi/abs/10.1080/10926480709336752

[13] Hassan, T. A., Hollander, S., Van Lent, L., & Tahoun, A. (2019). Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics*, *134*(4), 2135-2202. https://doi.org/10.1093/qje/qjz021

[14] Hibbert, P. (2013). Approaching reflexivity through reflection issues for critical management education. *Journal of Management Education*, *37*(6), 803–827. https://doi.org/10.1177/1052562912467757

[15] Johansson Falck, M., & Okonski, L. (2023). Procedure for identifying metaphorical scenes (PIMS): The case of spatial and abstract relations. *Metaphor and Symbol*, *38*(1), 1-22. https://doi.org/10.1080/10926488.2022.2062243

[16] Kearney, R. (2002). *On stories*. Psychology Press. https://doi.org/10.4324/9780203453483

[17] Lakoff, G., & Johnson, M. (1980). *Metaphors We Live By*. University of Chicago Press.

[18] Lakoff, G., & Johnson, M. (2008). *Metaphors We Live By*. University of Chicago Press.

[19] McGlone, M. S. (2007). What is the explanatory value of a conceptual metaphor? *Language & Communication*, *27*(2), 109-126. https://doi.org/10.1016/j.langcom.2006.02.016

[20] Nimehchisalem, V. (2023). *Unforgiven*. In Nimehchisalem, V., & Geng, H. (Eds.). *Tests & Us - A Collection of Real Stories* (Vol. 2). (pp. 11-14). Generis Publishing.

[21] Nimehchisalem, V., & Geng, H. (Eds.). (2023). *Tests & Us - A Collection of Real Stories* (Vol. 2). Generis Publishing.

[22] Qammar, A., Wang, H., Ding, J., Naouri, A., Daneshmand, M., & Ning, H. (2023). *Chatbots to ChatGPT in a cybersecurity space: Evolution, vulnerabilities, attacks, challenges, and future recommendations.* arXiv preprint arXiv:2306.09255. https://doi.org/10.48550/arXiv.2306.09255

[23] Reiss, M. V. (2023). *Testing the reliability of ChatGPT for text annotation and classification: A cautionary remark.* arXiv preprint arXiv:2304.11085. https://doi.org/10.48550/arXiv.2304.11085

[24] Schultz, P. L., & Quinn, A. S. (2014). Lights, camera, action! Learning about management with student-produced video assignments. *Journal of Management Education*, *38*(2), 234–258. https://doi.org/10.1177/1052562913488371

[25] Tan, Y., Min, D., Li, Y., Li, W., Hu, N., Chen, Y., & Qi, G. (2023, October). *Can ChatGPT replace traditional KBQA models? An in-depth analysis of the question answering performance of the GPT LLM family*. In International Semantic Web Conference (pp. 348-367). Springer Nature Switzerland.

[26] Taylor, S. S., & Statler, M. (2014). Material matters increasing emotional engagement in learning. *Journal of Management Education*, *38*(4), 586–607. https://doi.org/10.1177/1052562913489976

[27] ÜNLÜ, C. (2023). Interpretutor: Using large language models for interpreter assessment. In *Proceedings of the International Conference HiT-IT* (pp. 78-96).

[28] Zhu, D., Chen, J., Shen, X., Li, X., & Elhoseiny, M. (2023). *Minigpt-4: Enhancing vision-language understanding with advanced large language models*. arXiv preprint arXiv:2304.10592. https://doi.org/10.48550/arXiv.2304.10592

Hui Geng holds a Ph.D. in English Language from Universiti Putra Malaysia and is currently a senior lecturer at Guilin University of Technology, China. Her research interests encompass corpus linguistics, academic writing, EAP, ESP, discourse analysis, language assessment, creative writing, literature, and translation. She serves as the copy editor for the *International Journal of Education and Literacy Studies* (IJELS) and as the assistant editor for the *ASEAN Journal of Applied Linguistics* (ASJAL). She has translated two books, edited two books, and published several articles in SSCI- and Scopus-indexed journals. Email: huiyuqiankun@gmail.com

Han Wei holds a Ph.D. in English Language and Literature and has been actively engaged in English language teaching and research since the early 1990s. He served as a visiting scholar at the Department of Linguistic Studies, University of Texas at Arlington, U.S., in 2007, and at the Eastern Institute of Technology, New Zealand, in 1996. Since 2007, he has held a professorship at Guilin University of Electronic Technology, China. His research interests span corpus linguistics, translation studies, cognitive linguistics, and cross-cultural communication, with a particular focus on classical Chinese poetry composition and comparative linguistics between English and Zhuang. Email: 237523688@qq.com

Vahid Nimehchisalem holds a Ph.D. in TESL and started teaching English in 1994. He has since been involved in the area as a lecturer, teacher trainer, material developer, researcher, editor, and writer. He is currently a professor at the University of Economics and Human Sciences in Warsaw, Poland, co-director of the Australian International Academic Centre, and a former associate professor at Universiti Putra Malaysia. He is chief editor of the *International Journal of Education and Literacy Studies* (IJELS). His main area of teaching and research interest is research methods and language assessment. Emails: nimechie@gmail.com; v.nimehchisalem@vizja.pl

Ali Sorayyaei Azar holds a Ph.D. in Applied Linguistics from the University of Malaya and is currently an assistant professor at the Faculty of Education, University of Malaya, Malaysia. With a career spanning 27 years, he has worked as an EFL/ESL senior lecturer, researcher, and curriculum developer in Iran, Turkey, Qatar, and Malaysia. His research interests include academic discourse analysis, genre analysis, corpus linguistics, TESL, educational technology, and English language teaching strategies. He has published and presented research papers in several prestigious journals. He is the head of MAAL-SIG (Research Methods in Applied Linguistics) and serves as the editor for three books, all published by IGI Global Publisher. Additionally, he is a member of the Editorial Board for several academic journals. Email: azarsorrayaie@um.edu.my