

## Original Research

# Mind vs. machine: Comparative analysis of metaphor-related word translation by human and AI systems

by Zhengjian Li<sup>1</sup> and Lang Chen<sup>2</sup>

<sup>1</sup>Saarland University, Saarbrücken, Germany

<sup>2</sup>Guangdong University of Foreign Studies, Guangzhou, China

*The present study presents a comparative analysis of the translation processes and outcomes of human translators, Neural Machine Translation (NMT) systems and Large Language Models (LLMs) focusing on the translation of Metaphor-related Words (MRW). The study employs various research methodologies, including product analysis, think-aloud protocols, subsequent interviews, and translation quality assessments to uncover the choice of strategies in translating MRWs by different subject groups as well as its relation with quality criterion. Human translators and LLMs tend to favour strategies such as metaphor into different metaphor (M→M2) and metaphor reduction (M→Non), while NMT systems prefer the reproduction of metaphors (M→M). LLMs demonstrate translation patterns which are more aligned with human translators, helping them achieve higher evaluation scores, though their performance remains inconsistent, particularly with novel metaphors. Additionally, human translators process metaphors by incorporating conceptual, cultural, and contextual factors, whereas LLMs tend to rely on paraphrastic approaches. Evaluation results indicate that LLMs exhibit proficiency on par with novice translators in terms of accuracy, idiomatic expression, and vividness in metaphor translation, while NMT systems fall slightly short. The study highlights the influence of translation strategies on the quality of metaphor translation and concludes that, while NMT systems and LLMs can achieve performance comparable to human translators, much larger metaphor-specific datasets supported studies are expected to validate its consistency.*

**KEYWORDS:** machine translation, neural machine translation, metaphor translation, think-aloud protocol, translation strategy, large language model, computational translation, cognitive processing, conceptual metaphor, ChatGPT, DeepL

**CRedit AUTHOR STATEMENT:** Zhengjian Li: Methodology, Formal Analysis, Data Curation, Visualisation, Writing – Original Draft; Lang Chen: Methodology, Conceptualisation, Writing – Review & Editing, Supervision

**CONFLICT OF INTEREST:** The authors declared no conflict of interest

**DATA AVAILABILITY STATEMENT:** The data supporting this study's findings are available from the corresponding author, Lang Chen, at [clheather8803@sina.com](mailto:clheather8803@sina.com), upon reasonable request

**FUNDING:** No funding was reported for this research

**ARTICLE HISTORY:** Submitted September 27, 2024 | Revised January 10, 2025 | Accepted February 3, 2025

**FOR CITATION:** Li, Z., & Chen, L. (2025). Mind vs. machine: Comparative analysis of metaphor-related word translation by human and AI systems. *Training, Language and Culture*, 9(1), 10-27. <https://doi.org/10.22363/2521-442X-2025-9-1-10-27>



This is an open access article distributed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0), which allows its unrestricted use for non-commercial purposes, subject to attribution. The material can be shared/adapted for non-commercial purposes if you give appropriate credit, provide a link to the license, and indicate if changes were made.

## 1. INTRODUCTION

Metaphor, as a psychological and cognitive phenomenon instead of a pure linguistic figure of speech, is the process and product of cross-domain mapping within linguistics. As Schäffner (2004) indicates, 'metaphors are not just decorative elements, but rather, basic resources for thought processes in human society,' and the 'analysis of texts with respect to metaphors and metaphorical reasoning processes in different languages can, thus, reveal

possible cultural differences in conceptual structures' (Schäffner, 2004, p. 1258-1267). How to better utilise these 'basic resources' can be a focus for both human translators and machine translation systems in developing their translation potential. Metaphor translation, accordingly, involves complicated processes, such as re-conceptualisation and re-mapping, to produce authentic and idiomatic target texts, taking culture-specific elements into account (Massey & Ehrensberger-Dow, 2017).

Cognitive approaches to Translation Studies (TS) have a rich literature in reference, but the mechanism of conceptual metaphor working with TS has only emerged in recent years, relatively intermittently and insufficiently compared to metaphor research in a multifaceted way in cognitive linguistic-discourse studies and corpus linguistic studies (Demjén & Semino, 2016; Tendahl & Gibbs, 2008). Particularly focused on by researchers in TS is the investigation into the cognitive effort required in processing metaphors, which utilises varied experimental methods, including keylogging and eye-tracking (Jankowiak & Lehka-Paul, 2022; Massey & Ehrensberger-Dow, 2017; Sjørup, 2013). Optimally, the outcome of any found-out cognitive factors active within the process of TS are of great implicature for machine translation systems to better simulate human translation behaviour. While machine translation of metaphor is regarded as a downstream task of cognitive processing of metaphor and receives limited attention (Tong et al., 2021), the investigation of metaphor has become inevitable in NLP due to its ubiquitous presence in discourse (Steen et al., 2010). Despite research efforts to identify and extract the salient features of metaphorical expressions within natural language (Jang et al., 2016; Stowe et al., 2021), there is a paucity of literature specifically addressing the contribution of these features when working with machine translation. Mao et al. (2018) developed an unsupervised model to identify and paraphrase metaphors into plain texts and use them as input for machine translation systems. Even though the study suggests improved translation quality via human evaluation, metaphor is treated as an exception necessitating pre-processing before effective integration into machine translation systems.

The advent of large language models (LLMs) such as GPT-4 and LLaMA has prompted increased interest in investigating their capacity for metaphor processing, as a body of scholarly literature indicates that these models possess a certain level of metaphorical understanding (Aghazadeh et al., 2022; Ichien et al., 2023; Wachowiak & Gromann, 2023), given their demonstrated proficiency in metaphor detection, paraphrasing, and interpretation. However, it is also noted that these language models, which gain information through machine learning, operate *'in a way that is strangely different from how we commonly understand human intelligence,'* and their statistical structure makes the output *'difficult to explain'* (Dove & Fayard, 2020). Lin (2023) even criticises that *'the technology has become increasingly removed from the substance of language,'* as exemplified by the fact that these machines process 'informational' text more efficiently than 'expressive' text. Cai et al. (2024) subjected ChatGPT to 12 experiments to prove whether it can develop human-like characteristics in language use. Though these experiments prove that ChatGPT is capable of imitating human language processing, its internal mechanism remains a black box.

Existing research suggests that encoded knowledge within language models aids in metaphor processing beyond contextual information (Aghazadeh et al., 2022; Wachowiak & Gromann, 2023). Nevertheless, Neidlein et al. (2020) propose that this

capability primarily stems from *'optimising the disambiguation of conventionalised, metaphoric word senses for specific words instead of modelling general properties of metaphors'* (Neidlein et al., 2020, p. 3722). According to this perspective, the capability of language models in metaphor processing is a by-product based on their deep understanding of the semantic meanings of words. Under the ongoing transition from machine translation to LLMs, metaphor becomes a crucial tool to explore whether these models can truly process the implicit meaning of language; little research has been done on the comparison between human translators and language models on the metaphor translation task from a cognitive perspective. Much beyond, the use of metaphor is highly hypothesised to be varied by the nature that the embodiment of metaphor embraces cognitive flexibility and cultural diversity (Littlemore, 2019), accordingly, translation process of metaphor as well as product could be more or less unstable or inconsistent despite this study is devoted to search for the patterns and traits among human translators, neural machine translation (NMT) systems, and large language models (LLMs). With combined methods of product analysis with translation quality assessment and process-oriented approach such as think-aloud protocols, human translators' cognitive process when translate metaphor are revealed and the contrasts between human translators and AI systems are straightened out.

## 2. THEORETICAL BACKGROUND

### 2.1. Metaphor translation at linguistic, cultural and cognitive levels

Metaphor, since it was first proposed by Lakoff and Johnson (1980) with an extended connotation, can be considered as not only an isolated figurative expression but a process of cross-domain mapping between two distinct concepts, which creates metaphorical expressions to convey abstract concepts in the target domain (Massey, 2021; Thibodeau et al., 2019). Metaphor translation, as *'the most important particular problem,'* is tortuous for its challenge to a translator *'at least doubled – if not squared – when two languages come into play'* (Rojo, 2015, p. 721).

In traditional perspectives, metaphor translation studies have been centrally concerned with *'firstly, the translatability of metaphors, and secondly, the elaboration of potential translation procedures'* (Schäffner, 2004, p. 1253). The viewpoints on the translatability of metaphors are divergent: some scholars posit that any attempt to translate a metaphor results in a new metaphor in the target language (Dagut, 1987; Nida, 1964; Vinay & Darbelnet, 1995); others support the full translatability of metaphor (Kloepfer, 1981; Mason, 1982). As proposed by Kloepfer (1981), metaphor translation does not present a challenge to the translator since humans share the same imaginary. However, some scholars also consider that metaphors are quasi-translatable, i.e. they are translatable under some constraints (Newmark, 1981; Snell-Hornby, 1988; Toury, 1985; Toury, 2012; van den Broeck, 1981). Cultural disparities are poised as a focus for researchers to determine if a metaphor is translatable. Mason (1982) aligns with this notion, acknowledging that cultural

disparities can render some metaphors amenable to direct translation while necessitating alternative strategies for others. Based on these viewpoints, a set of common metaphor translation procedures is established by various scholars (Newmark, 1981; Toury, 2012; van den Broeck, 1981). For instance, Toury (2012) concludes six metaphor translation strategies from both the source and target poles, viz. metaphor into the same metaphor (M→M), metaphor into a different metaphor (M→M2), metaphor into non-metaphor (M→Non), metaphor into nothing (M→∅), non-metaphor into a metaphor (Non→M), nothing into a metaphor (∅→M). Toury's (2012) methodology liberates metaphor translation studies from the singular focus on the source text. By integrating both the source and target texts into a unified framework, it enriches the research on metaphor translation. The vast majority of research findings have elevated metaphor translation beyond its conventional linguistic realm.

Even though the perspectives on metaphor translation have expanded from a linguistic to a cultural level, these non-cognitive approaches are not able to unveil a deeper layer of metaphor translation, i.e., cross-domain mapping. Numerous studies thus incorporate human cognition as a pivotal element in investigating metaphor translation, thereby bridging the gap between linguistics and cognitive processing. Shuttleworth (2011) analyses 1354 metaphor examples and their officially published translations in five different languages, scrutinising the levels of both individual metaphorical expression and metaphorical mapping, which reveals that while explicit renderings exist at the linguistic level, the metaphorical mappings tend to be preserved. Notwithstanding that many scholars try to unveil the mapping mechanism through corpus-based approaches and comparison analyses (Schäffner, 2004), the cognitive process in metaphor translation is still hard to trace back through these product-oriented approaches. Experimental approaches, including keylogging and eye-tracking, are thus adapted in metaphor translation research. Massey and Ehrensberger-Dow (2017), for instance, analyse translation processes and output data from target-text products, keystroke logs, and retrospective verbal commentaries. The findings suggest that when human translators approached the translation of two complex conceptual metaphors from German to English, complex metaphors may possess a degree of cultural specificity and indicate that the practices of re-mapping in translation are influenced by the experience of the translator. Additionally, the results demonstrate that re-mapping to a source-language target domain may generate more uncertainty compared to generic-level re-mapping. By combining process-oriented and product-oriented research methods, these studies reveal metaphor translation mechanisms from a deeper perspective, including cognitive effort and translation decision-making.

It is true that metaphors often carry strong cultural and contextual connotations with its understanding, transferring, and producing in translation which requires cognitive efforts to capture the nuances of accurate information. This poses a challenge not only for human translators but also remains as a

primary task for the progress of machine translation. Koglin (2015) investigates the cognitive demand on translators in the post-editing of machine-translated metaphors compared to that in manual metaphor translation, with eye tracking and keystroke logging as measurement tools, and the research reveals that manual metaphor translation requires higher cognitive effort. Koglin and Cunha (2019) continue to develop the topic of 'cognitive effort' in post-editing metaphors translated by two different machine translation systems, elucidating that the outputs from the hybrid machine translation system consume higher cognitive effort in the post-editing task than those from the statistical machine translation system. Even though it cannot be excluded that other factors, such as the loss of contextual information stemming from low-quality MT outputs, may influence the consumption of cognitive effort, the two experiments propose a creative methodology grounded in cognitive effort for detecting metaphor translation quality.

Different from the aforementioned studies that focus mainly on post-editing translation, Massey (2021) expanded upon their previous study by conducting a comparison between human translators and NMT systems in their handling of two complex culturally specific conceptual metaphors during the translation process. This comparison illustrates that what distinguishes experienced translators from publicly available NMT systems is their ability to access the conceptual level of meaning realised in a particular lexical form, and then to generate multiple potential target-text solutions to reach a given target audience. Additionally, the decision-making and problem-solving of human translators occur on a conceptual level, regardless of their training and experience levels. Although the study contributes significantly to metaphor translation and machine translation, the data analysed in this study are not specifically for researching conceptual metaphor, which may pose challenges in the interpretation of the results.

The previous research reveals the complexity of metaphor translation, which entails languages, cultures, cognitive processes, and other factors. Nevertheless, the focal point predominantly centres around the comparisons between source and target translations while the investigation of translators and their translating processes remains rare.

## 2.2. Metaphor translation at computational level: from NMT to LLMs

In the realm of linguistic computational models, the capabilities of NMT and LLMs in facilitating translation tasks have been substantiated through empirical researches. Junczys-Dowmunt et al. (2016) have elucidated the efficacy of NMT, positing that its proficiency in terms of words-per-second ratios renders it a viable option for deployment within in-production systems. Subsequently, Kocmi and Federmann (2023) have explored the utility of pre-trained, generative LLMs in the evaluation of translation quality, thereby offering preliminary insights into their potential applications in ensuring the fidelity of translations. Metaphor translation, when performed by language

models, involves the intersection of metaphor identification and metaphor interpretation, both of which are sub-tasks within the field of computational metaphor processing.

Metaphor identification aims to detect if a given text by different semantic units, e.g., words, phrases, and sentences, is labelled as being used metaphorically or literally, which is often considered a binary classification task in a considerable number of studies. Do Dinh and Gurevych (2016) present a neural network to detect metaphor in conjunction with pre-trained word embeddings from a token level, achieving positive results. The metaphor identification task is also adopted by LLMs in order to test their performance in different downstream tasks. Wachowiak and Gromann (2023) investigate the ability of GPT-3 to detect conceptual metaphors and predict their source domains, yielding an accuracy of 60.22% in the prediction task. Tong et al. (2024) released a dataset called the Metaphor Understanding Challenge Dataset specific for testing how well LLMs understand metaphor, and their experiments with LLaMA and GPT-3.5 demonstrate that enhanced performance is observed in the models when the metaphorical word is asked to focus on, and the novelty of the metaphorical word may affect their performances.

According to Ge et al. (2023), property extraction, word-level paraphrasing, and explanation pairing are three different approaches to metaphor interpretation. Property extraction interprets metaphors by extracting shared features from source

and target domains, while word-level paraphrasing involves the substitution of a metaphor with its literal words, and explanation pairing appends a literal explanation to a metaphorical expression. Additionally, generating metaphors with a literal input can also be considered as an approach to metaphor interpretation, although in a reversed direction. For example, Stowe et al. (2021) fine-tuned a pretrained BART model with conceptual mappings between cognitive domains encoded to generate metaphors, which shows state-of-the-art performance by both automatic and human evaluations. Tong et al. (2021) indicate that the models are likely generating the word without 'knowing' that it is a metaphor and highlight that '*it is thus in the intentional use of metaphors that the difficulty of metaphor generation lies.*'

Advancement of AI systems have consistently demonstrated notable capabilities in natural language generation. However, their performance in the domain of machine translation remains underexplored, with comprehensive process and product analyses yet to be conducted (Hendy et al., 2023).

### 3. MATERIAL AND METHODS

In the subsequent study, efforts are made to compare translations provided by human translators and language models. The study incorporated data from translation assessments, think-aloud protocols, and subsequent interviews. This section will outline the procedure and research method used (Figure 1).

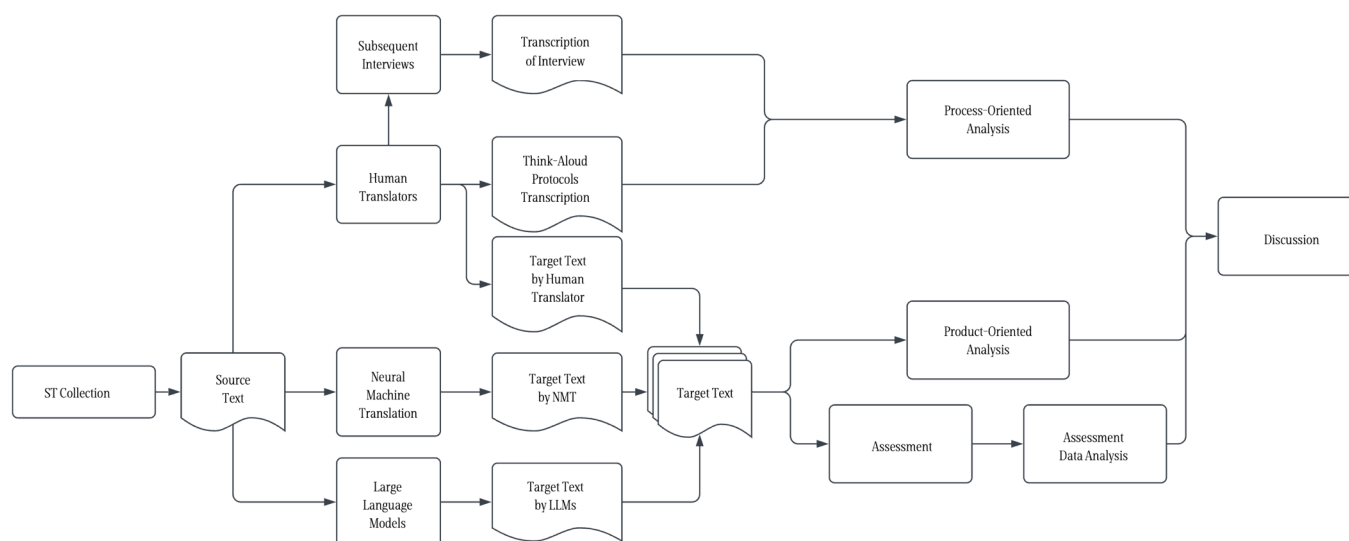


Figure 1. Flowchart of the study phases

The study selected 30 groups of English sentences from 30 articles respectively, containing MRWs identified with MIPVU as the source texts (STs) for further research and analysis. Each group contains one MRW to be investigated. Most groups consist of a single sentence, while others contain two or more sentences to offer sufficient context for translation. The STs are randomly sourced from 30 articles between April 2022 and February 2023 on topics about international politics, economics

and business, science and technology, climate and environment, and arts and lifestyle from *The New York Times*. The MRWs, which can be verbs, nouns, or adjectives, are underlined for investigation.

The term '*metaphor-related word (MRW)*' is defined as '*a word is used indirectly, and that use may potentially be explained by some form of cross-domain mapping from a more basic meaning of that word*' (Steen et al., 2010, p. 766).

Furthermore, metaphoricity is dynamic and contains a spectrum of gradations, rather than adhering to a binary classification (Hanks, 2006). This indicates that a metaphor can become conventionalised when it is frequently used in human society, thereby evolving into a familiar expression (Croft & Cruse, 2004) or feel creative with a sense of originality yet appropriateness fulfilling communicative purpose (Littlemore & Fielden-Burns, 2023). Steen et al. (2010) extend the Metaphor Identification Procedure (MIP) to the Metaphor Identification Procedure Vrije Universiteit (MIPVU), based on a previous study by Pragglejaz Group (2007), to identify metaphors and classify metaphor-related words (MRWs) into different categories, namely, MRW indirect, MRW direct, MRW implicit, and MFlag, according to their metaphoricity. The MRWs from the current study mostly fall into the category of MRW direct, which is characterised by a high degree of metaphoricity, making them challenging to recognise. The study also includes instances of MRW indirect and MRW implicit; however, MFlag is not represented.

The fundamental hypothesis of MIP, i.e., a MRW conveys specific contextual meaning distinct from its basic one, performs effectively in language models for computational metaphor identification (Song et al., 2021; Zhang & Liu, 2022; Lin et al., 2021). However, the hypothesis falls short in certain situations (Choi et al., 2021; Mao et al., 2019).

In the qualitative study, 5 MA (Master of Arts in Translation and Interpreting: Academic Degree Program)/MTI (Master of Translation and Interpreting: Professional Degree Programme) graduate students (MAG) and 5 BA (Bachelor of Arts) students (BAG) who were Chinese native speakers with English-Chinese CATTI Certificate Level-3 or higher (CATTI: China Accreditation Test for Translators and Interpreters), participated in translating the English STs into the Chinese target texts (TTs). Each participant translated 6 groups of the STs, and 60 groups of TTs were ultimately collected. Notably, each group of STs was translated by two translators, one from MAG and one from BAG. The study was conducted remotely using Tencent

Meeting, allowing the participants to work at their own pace. During the translation task, they were permitted to utilise electronic dictionaries and search engines but not machine translation as well as generative language models. During the process of translation, they were required to verbalise their thought processes as the immediate think-aloud protocols (TAPs). A brief training was provided to guide them to verbalise their thinking processes before the task began, and both Chinese and English were allowed to be used during the verbalisation process. After the translation task, participants were asked to do subsequent interviews which required them to analyse and review their translations. The participants' translation and verbalisation processes, as well as subsequent interviews focusing on MRWs, were recorded and a verbatim transcription of the recordings was manually documented.

The outputs of machine translation were collected in February 2023 by manually inputting each syntagma of the ST into both DeepL Translator (a neural machine translation service which claims to be the world's most accurate and nuanced translator, hereafter referred to as DeepL), which offers high-quality translations, and Youdao Fanyi (a popular translation service developed by the Chinese local company NetEase, hereafter referred to as Youdao). Both machine translation systems utilise NMT models. ChatGPT, as a large language model, relies on prompts as commands to accomplish specific tasks, including translation, and the style of prompts influences its output.

In an examination of ChatGPT's translation capabilities, Jiao et al. (2023) assessed the performance of three prompts for Chinese-to-English translation tasks, identifying one prompt that outperformed the others in terms of BLEU, ChrF++, and TER metrics.

This study adheres to the translation prompt which is: 'Please provide the Chinese translation for these sentences.' The translations of Large Language Models were collected in December 2024 by manually inputting each syntagma with the prompt of the ST into ChatGPT-3.5 (hereafter GPT-3.5) and ChatGPT-4 (hereafter GPT-4).

Table 1  
*Translation strategies*

ST-based	1	M→M	metaphor into same metaphor
	2	M→M2	metaphor into different metaphor
	3	M→Non	metaphor into non-metaphor
	4	M→∅	metaphor into nothing
TT-based	5	Non→M	non-metaphor into metaphor
	6	∅→M	nothing into metaphor

*'Junczys-Dowmunt et al. (2016) have elucidated the efficacy of NMT (Neural Machine Translation), positing that its proficiency in terms of words-per-second ratios renders it a viable option for deployment within in-production systems. Subsequently, Kocmi and Federmann (2023) have explored the utility of pre-trained, generative LLMs (Large Language Models) in the evaluation of translation quality, thereby offering preliminary insights into their potential applications in ensuring the fidelity of translations. Metaphor translation, when performed by language models, involves the intersection of metaphor identification and metaphor interpretation, both of which are sub-tasks within the field of computational metaphor processing'*

The MRWs of TT were categorised following the taxonomy outlined in Table 2, an adaptation from Toury's (2012) strategies by Nacey and Skogmo (2021). As for the present study, the TT-based strategies (categories 5 and 6 in Table 2) handling metaphor translation are deemed impractical, as they are applicable only in cases where no metaphor is present in the text. Since the focus is how metaphors are translated, four source text-based strategies are employed.

To evaluate the quality of translations of texts that contain metaphors, assessment criterion includes accuracy, idiomaticity, vividness, and acceptability. Accuracy and idiomaticity are the most commonly used criteria to evaluate translations, as they are essential for ensuring that the intended meaning of the source text is conveyed precisely and naturally in the target language. In particular, accuracy requires the translator to convey the meaning of MRWs as closely as possible to the original text, while idiomaticity requires the translation to read as though it were originally written in the target language (Ghazala, 2003; Hale & Campbell, 2002). Vividness is a dimension that is especially relevant to translations that contain MRWs, as metaphors are often used to produce aesthetic effects in language. As the same as idiomaticity, vividness shows the translator's ability to convey the meaning of the MRWs in a creative and impactful way, using language that emotionally resonates with the reader and leaves a lasting impression. Achieving vividness in translation requires the translator to employ a range of literary techniques and to be highly attuned to the nuances of the target language. Acceptability, first proposed by Toury (2012), indicates the extent to which the translation is understood and accepted by its target audience. It reflects the fact that translations are created for readers, and their perception of the translation is a crucial determinant of its success or failure. To some extent, it closely resembles the aforementioned 'appropriateness' at the operational level. In sum, it is essential to ensure that translations of texts containing MRWs are accurate, idiomatic, vivid, and acceptable to their readership.

Different from the human-based metric in this study, automatic metrics are also used for machine translation assessment. Word overlap-based metrics, including BLEU (Papineni et al., 2002), ROUGE (Lin, 2004), METEOR (Banerjee & Lavie, 2005), broadly used in evaluation, are not ideal for machine translation. For example, Tran et al. (2019) demonstrate that the weak correlation between BLEU and the semantic correctness diminishes BLEU's effectiveness in evaluating translation quality. In recent years, model-based metrics like Sentence Movers Similarity (Clark et al., 2019), BLEURT (Sellam et al., 2020) alternatives, and LLMs-based metrics like G-Eval (Liu et al., 2023) and LLM-as-a-judge (Zheng et al., 2023) are also used to evaluate generation outputs.

Even though these automatic metrics provide standardised and comparable results, the present study, aiming to conduct a comparison from a human perspective, is better suited to a human-based assessment against the background that the cognitive mechanism of human translators represents a key value-added component for disembodied artificial intelligence (Massey, 2021).

To assess translations on the four identified dimensions, an evaluation form (Figure 2) was developed and administered to both machine and human translations. The form utilised a Likert Scale, with assessors asked to assign a score ranging from 1 (indicating strong disagreement) to 5 (indicating strong agreement) based on the provided statements. For instance, an assessor was asked to respond to the scale according to a statement like 'the translation of the underlined word is accurate'.

A total of 10 assessors, all of whom were excellent BA or MA/MTI students who had previously won awards in translation competitions like the Han Suyin International Translation Contest (the longest-running, largest-scale and the most influential high-level professional translation event in China), were invited to evaluate the translations. Notably, these assessors were distinct from the translation participants and they were specifically chosen for the evaluation tasks. Each assessor was responsible for evaluating 12 translations. Four of the ten assessors were re-recruited to assess another 60 translations generated by LLMs, with each assessing 12 translations. They are well trained beforehand according to the assessment criterion with discussion to ensure the inter-related reliability.

This study adhered to ethical principles to ensure the protection of participants' rights and privacy. Before the study commenced, all participants signed an informed consent form, confirming their full understanding of the research objectives, procedures, data usage, and their right to withdraw at any time without any negative consequences. To safeguard participants' confidentiality, all data were anonymised and used solely for research purposes. Participation in this study was entirely voluntary, and participants were free to withdraw at any stage. As a token of appreciation for their time and effort, each participant received a compensation of 100 RMB. The study was conducted in accordance with established ethical guidelines for academic research.

On a scale of 1 to 5, where 5 indicates strong agreement and 1 indicates strong disagreement, lower scores represent lower levels of agreement.

The translation of the underlined word is accurate	1	2	3	4	5
The translation of the underlined word is idiomatic	1	2	3	4	5
The translation of the underlined word is vivid	1	2	3	4	5
The translation of the underlined word is acceptable	1	2	3	4	5

Figure 2. Sample assessment form

## 4. STUDY RESULTS

### 4.1. Product-oriented analyses

#### 4.1.1. Pattern of translation strategies

In this study, the translation task was administered to a cohort of 30 sets of sentences containing MRWs, which were subjected to translation by six distinct entities: two human translation (HT) groups, identified as BAG and MAG; two neural machine translation (NMT) systems, namely, DeepL and Youdao; and two large language models (LLM), specifically, GPT-3.5 and GPT-4. This yielded a total of 180 sets of target texts (TTs), with each translating entity contributing 30 sets of TTs. Consequently, the translation output was segmented into 60 sets of TTs for each of the three categories: HT, NMT, and LLM, with each category further subdivided into 30 sets per translating entity. The strategies employed by different groups in their translations are detailed in Table 2 that provides a comprehensive comparison of translation strategies across several groups, delineating the frequency of different approaches to MRWs. The comparison between the NMT group and HT group shows a

marked difference in the translation strategies chosen for MRWs. While the NMT group tends to adhere more closely to the original metaphor applying literal translation (M→M), the HT group exhibits a more liberal approach with greater instances of paraphrasing the meaning of the MRWs (M→Non). Additionally, the HT group demonstrates a higher rate of translating the metaphor into another metaphor (M→M2) than that of the NMT group. The LLM group exhibits a comparable pattern to the HT group as its primary strategy. Within the NMT group, DeepL exhibits a pronounced preference for maintaining the original metaphor (M→M). Youdao, on the other hand, demonstrates a more diversified strategy application, engaging equally in M→M2 and M→Non strategies. BAG and MAG, being groups of human translators, tend to use M→Non as the primary strategy, but BAG has a relatively high percentage of M→M2, while MAG has a lower percentage. In the LLM group, GPT-3.5 and GPT-4 showcase a close alignment in ratios across different strategies, demonstrating a significant level of consistency in their translation patterns.

Table 2  
Overview of translation strategies used in different groups

	M→M	M→M2	M→Non	M→∅
NMT (N = 60)	29 (48.3%)	10 (16.7%)	20 (33.3%)	1 (1.7%)
↳ DeepL (n = 30)	21 (70.0%)	4 (13.3%)	5 (16.7%)	–
↳ Youdao (n = 30)	8 (26.7%)	6 (20.0%)	15 (50.0%)	1 (3.3%)
HT (N = 60)	5 (8.3%)	19 (31.6%)	32 (53.3%)	4 (6.7%)
↳ BAG (n = 30)	2 (6.7%)	13 (43.3%)	13 (43.3%)	2 (6.7%)
↳ MAG (n = 30)	3 (10.0%)	6 (20.0%)	19 (63.3%)	2 (6.7%)
LLM (N = 60)	15 (25.0%)	12 (20.0%)	33 (55.0%)	–
↳ GPT-3.5 (n = 30)	8 (26.7%)	6 (20.0%)	16 (53.3%)	–
↳ GPT-4 (n = 30)	7 (23.3%)	6 (20.0%)	17 (56.7%)	–

#### 4.1.2. MRW translation based on metaphoricity

This section considers in detail the practical use of translation strategies for MRWs by presenting a detailed case study.

Various types of metaphors, ranging from conventional to novel, are examined to show how MRWs are adapted from the source language to the target language.

Table 3

Examples with different metaphoricity degrees and adopted translation strategies (MRWs are underlined)

CATEGORY	CONTENT	STRATEGY
ST	(1) 'I was really starting to feel financially <u>squeezed</u> in a way that I hadn't felt ever before, since finishing college,' Mr. Bate said.	
DeepL	贝特先生说: '我真的开始感到经济拮据, 这是我在大学毕业后从未感受过的。'	M→Non
Youdao	我真的开始感到经济拮据, 这是我大学毕业以来从未有过的感觉, '贝特说。	M→Non
BAG	贝特先生说: '我确实开始感到经济拮据, 在大学毕业后就从来都没有过这样的感觉了。'	M→Non
MAG	贝特先生说: '从大学毕业之后, 我真的开始感觉手头上很拮据, 这是我之前不曾感受到的。'	M→Non
GPT-3.5	我真的开始感到经济上的压力, 这是我大学毕业后从未有过的感觉。' 贝特先生说道。	M→M2
GPT-4	'自从大学毕业以来, 我真的开始感到在财务上被挤压, 这是我以前从未有过的感觉, '贝特先生说。	M→M
ST	(2) The Biden administration is trying to use uncommon legal theories to <u>clip the wings</u> of some of America's largest businesses.	
DeepL	拜登政府正试图用不常见的法律理论来剪掉美国一些最大企业的翅膀。	M→M
Youdao	拜登政府正试图利用不同寻常的法律理论来削弱美国一些大型企业的实力。	M→Non
BAG	拜登政府如今正试图借由冷门的法律理论来打压美国的一些商业巨头。	M→M2
MAG	拜登政府企图倚靠不寻常的法学理论去打压美国最大的某些公司。	M→M2
GPT-3.5	拜登政府正在尝试使用不常见的法律理论来削弱一些美国最大企业的影响力。	M→Non
GPT-4	拜登政府正尝试使用不常见的法律理论来限制一些美国最大企业的权力。	M→Non
ST	(3) It has its diplomatic <u>eggs</u> in many baskets, and its president said he wasn't concerned about pressure from rich, industrialised countries to drop the campaign for an international legal opinion.	
DeepL	它的外交鸡蛋放在许多篮子里, 它的总统说他并不担心来自富裕的工业化国家的压力, 以放弃争取国际法律意见的运动。	M→M
Youdao	它在外交上有很多选择, 它的主席说, 他并不担心来自富裕工业化国家的压力, 要求他们放弃争取国际法律意见的运动。	M→Non
BAG	国家对很多事件都有特定的外交方针政策, 并且国家主席说他并不担心富裕的工业化国因为一个国际化的合法观点发起反对运动, 向自己国家施压。	M→Non
MAG	这个国家外交资源充足, 总统也表示并不在乎富裕的工业化国家联合起来发表法律意见书。	M→Non
GPT-3.5	它在许多篮子里放置了它的外交'鸡蛋', 而它的总统表示, 他并不担心来自富裕工业化国家的压力, 要求放弃国际法律意见的运动。	M→M
GPT-4	它在许多领域进行了外交投资, 其总统表示他并不担心来自富裕工业化国家的压力, 要求放弃寻求国际法律意见的运动。	M→Non

Three representative examples are all direct metaphor but demonstrate varying degrees of metaphoricity from conventional to novel (Examples 1 to 3 in Table 3). They are provided to demonstrate how metaphor-related words (MRWs) are translated from the source language to the target language approached by different groups. The findings include the following.

1. The complexity arising from varying degrees of metaphoricity has a direct impact on the variability in metaphor translation strategy choices. But in this part, it's yet to be decisive in determining the patterns and extent of variability as well as the consistency across human translators, neural machine translation, and large language models, since measurement of metaphoricity requires another rigorous experiment so that establish correlations and causality with the six metaphor translation strategies with a further step, which we will discuss in the other article. The focus of the statistics in the current study is to explore the comparative and contrastive performance regarding the accuracy, idomaticity, vividness, and acceptability of translations produced by human translators and various AI systems.

2. Despite that, the descriptive statistical results of metaphor translation strategies as chosen by human translators and LLMs as shown in Table 2 demonstrate the potential similarities between the two in terms of cognitive reasoning and decision-making. This finding corroborates Brown et al. (2020) that the unique aspect of GPT models lies in their ability to generate context-aware and coherent text. To illustrate, the preference of human translators and large models for the M→Non strategy aligns with the findings that GPT translations exhibit significant 'paraphrastic' characteristics, using one's own words to express. It is suggested that this approach could be applied to the handling of figurative text by neural machine translation models.

3. NMT systems tend to adhere closely to the superficial syntactic structure during the translation process and GPT-4's performance is better than GPT-3 and NMT systems (as illustrated in Table 2 and Example 2). NMT systems are inclined to

adhere to the superficial structure in the process of translation, whereas human translators and GPT may opt for explicit expressions in target texts.

#### 4.2. Process-oriented analyses

The immediate think-aloud protocols and the interviews are transcribed verbatim to facilitate detailed analysis. The transcripts reveal the thinking processes engaged in translation as well as the explanation of the strategies and techniques selected by the translators. Transcript analysis, with a particular focus on the translator's processing of MRWs, is demonstrated as follows.

##### 4.2.1. Processing MRWs

Despite not being provided with any information about MRWs in the translation process, the translators' understanding of MRWs can be deduced based on Case 1 and 2 as examples: 1) the process of understanding MRWs is highly automated; 2) MRWs can trigger a 'mental' image; 3) translators can identify novel MRWs by excavating their characteristics. In Case 1, during the translation of the MRW 'squeezed,' two translators immediately understood the intended meaning of the MRW. In their subsequent interviews, they mentioned that the MRW evoked a mental image in their minds, which was associated with the actual meaning. Case 2 demonstrates how translators understand novel MRWs that they haven't encountered before. From Case 2, it is evident that translators identify the intended meanings of novel MRWs by analysing their characteristics. Two translators successfully recognised 'oxygen' as an MRW and interpreted its intended semantics based on the attributes associated with oxygen. Both translators considered oxygen to be an extremely important element. Based on the preceding discussion, it can be concluded that during the translation process, translators initially grasp an abstract understanding of MRWs, then encode their overt connotations, and finally decode their covert implications through a cognitive mapping mechanism before deciding on the appropriate translation approach to use.

Table 4  
 Case 1 and 2 for case study (MRWs are underlined)

	ST	'I was really starting to feel financially <u>squeezed</u> in a way that I hadn't felt ever before, since finishing college,' Mr. Bate said.
CASE 1	TAPs transcript	Translator 1: '呃，首先来看主句就是 I was really starting to feel financially squeezed. 我确实感到，我确实开始感到经济拮据了。' ('Uh, first, let's look at the main sentence, which is 'I was really starting to feel financially squeezed.' I indeed felt, I indeed began to feel financially constrained.) Translator 2: 'Financially squeezed 经济上，感觉是经济上遇到一些困难。经济上遇到一些困难。' ('Financially squeezed, in terms of finance, it feels like encountering some difficulties financially. Encountering some difficulties financially.')
	Interview transcript	Translator 1: '对这个词的理解就是它其实特别的形象，squeeze，本来它是指那个，啊，抓紧了或者挤出来的意思。[...]就是就是一分钱都挤不出来了，那种感觉，首先就是它特别的形象。就是可以让人自然的联想到就是指经济拮据，没有钱了。' ('The understanding of this term is that it is actually very vivid. 'Squeeze' originally refers to, ah, the meaning of grasping tightly or squeezing out. [...] It is like not being able to squeeze out even a penny, that kind of feeling. First of all, it is particularly vivid. It naturally leads one to associate it with being financially strapped, having no money left.')

	Translator 2: ‘它本意就是挤压那个意思。呃，我们说很挤，就是在外面人群很挤的时候也会说squeeze。就有一点那种就是很挤，很挤，就是反映在financially，就是很紧张，那种感觉就没有太褒义，肯定不是说很有钱啊啥的。就感觉像手头挤出点钱出来这样的感觉。’ (‘Its original meaning pertains to compression or squeezing. When we say it’s very crowded, for instance, in a crowded outdoor setting, we also use the term ‘squeeze.’ It conveys a sense of being tightly packed, which, when applied to financial contexts, signifies a state of tension or strain. This connotation is certainly not positive, and it definitely does not imply affluence. It evokes the feeling of having to squeeze out some money from an already tight budget.’)
	ST During much of the 20th century until his death in 1965, the Swiss-French Modernist Le Corbusier sucked up most of the architectural oxygen in France, but there were other talents, some major, among them Jean Dubuisson (1914-2011).
	TAPs transcript Translator 1: ‘我结合上下文一看就会就会感受到，他会想去表达这个现代主义家很厉害的一种感觉。’ (‘Upon examining the context, one would immediately sense that it aims to convey a certain prowess associated with modernist figures.’)
CASE 2	Interview transcript Translator 1: ‘它肯定不会是氧气的意思[...] 氧气在我看来是一种很关键的东西吧，它既然它自己一个人能吸入那么多氧气，就说明它在这个方面还是比较重要，比较厉害的，[...]而且联系实际来想的话，这个地方他肯定是用到了一个比喻。’ (‘It definitely doesn’t mean oxygen [...] In my view, oxygen is quite crucial. Given that it can inhale so much oxygen on its own, this indicates that it is relatively important and formidable in this aspect. [...] Moreover, when considering the practical implications, this place must have employed a metaphor.’) Translator 2: ‘oxygen本意就是氧气，[...]然后oxygen再加上他本来就是，我们人活着必不可少的，然后也非常重要的东西，我就把它具象成了，就把引申成了知识。或者是我觉得这里翻译成养料养分也可以。’ (‘The original meaning of oxygen is indeed oxygen, [...] and then, considering that oxygen is essential and very important for human survival, I have concretised it, extending it to mean knowledge. Or I think translating it as nutrients could also be appropriate.’)

#### 4.2.2. Contextual factors

Translation of MRWs is significantly impacted by contextual factors. As shown in Cases 3-4, such factors facilitate the comprehension of MRWs. Taking into account various contextual factors, such as the subject matter and the stylistic features,

it is demonstrated through their TAPs in Case 4 and 5 that different translation strategies are employed based on situational demands and circumstances. In consideration of the aforementioned transcripts, it is evident that contextual information significantly influences the comprehension and processing of MRWs.

Table 5

Case 3, 4 and 5 for case study (MRWs are underlined)

	ST ‘I was really starting to feel financially <u>squeezed</u> in a way that I hadn’t felt ever before, since finishing college,’ Mr. Bate said.
CASE 3	TAPs transcript t Translator 1: ‘而且这个翻译[支持者]更加贴近这一个正式的语境。’ (‘Moreover, this translation [supporter] is more aligned with the formal context.’) Translator 2: ‘这个词[big fans]原本是粉丝的意思，然后就结合了一下这个语境，然后自己再翻译了一下这个词。’ (‘The term [big fans] originally means ‘fan’, and then, taking into account the context, I translated the term anew.’)
	ST The death toll in protests that have <u>convulsed</u> Peru rose above 20 on Friday, with tensions centred in the highland city of Ayacucho, where eight people were killed in clashes between anti-government demonstrators and military officers.
CASE 4	TAPs transcript t Translator 1: ‘还是在想convulse那个单词的意思。这个语境要怎么翻译才比较合适?’ (‘I’m still pondering the meaning of the word ‘convulse’. How should it be translated in this context to be more appropriate?’)
	Interview transcript t Translator 2: ‘呃，我看到这个词。首先也是跟上一题一样，就是根据一个语境。第一遍读的时候先去猜测它是一个什么样的意思。’ (‘Ah, when I encountered this term, it was similar to the previous instance, based on the context. Upon first reading, one would initially speculate about its meaning.’)

CASE 5	ST	Historical data has always been critical to those who make economic predictions. But three years into the pandemic, America is suffering through an economic <u>whiplash</u> of sorts — and the past is proving anything but a reliable guide.
	TAPs transcript	Translator 1: ‘根据语境来选择相关的词。再来考虑怎么翻译。’ (‘Select relevant words based on the context. Then consider how to translate.’)

4.2.3. Cultural awareness

The related work section elucidated that the social turn emphasised the profound relevance of cultural factors in metaphor and related investigations. In the present study, translators were observed to evince a keen regard for cultural factors shown in Case 6, 7, and 8.

By analysing the cases at hand, it becomes evident that translators exhibit a remarkable ability to link different cultural factors with one another. This is especially apparent in instances where they consider current political circumstances in their work, taking into account the unique cultural nuances that shape such events. Additionally, translators regularly consider cross-cultural factors such as linguistic differences and variations

in social norms when developing their translations. As a result, they are able to produce translations that accurately and effectively convey not only the literal meaning of the original text but also its cultural context and implications. The process of human translation transcends surface rendering and requires capturing the concept of a word, phrase (sentence), and the general idea of the whole message (text). As analysed in this section, in the process of translating MRWs, translators engage in a cognitive process of conceptual comprehension by analysing the overt and covert connotations of the MRW while considering various contextual factors. The translators also take into consideration the cultural background as it aids in understanding the meanings of the MRW and selecting the appropriate translation strategies.

Table 6  
Case 6, 7 and 9 for case study (MRWs are underlined)

CASE 6	ST	Historical data has always been critical to those who make economic predictions. But three years into the pandemic, America is suffering through an economic <u>whiplash</u> of sorts — and the past is proving anything but a reliable guide.
	TAPs transcript	Translator 2: ‘因为whiplash翻译成鞭子，鞭打这种意向，嗯，不管在什么文化里面应该都是可以理解到它是，有点负面的那种感觉的，所以就算是直译出来他鞭打也没有关系。’ (‘Because translating ‘whiplash’ as ‘whip’ or ‘whipping’ carries a connotation, well, in any culture, it should be understood to have a somewhat negative feeling, so even if it is literally translated as ‘whipping’, it would be acceptable.’)
CASE 7	ST	U.S. officials and analysts worry that Palestinian frustration with a <u>moribund</u> peace process, along with the rise of hard-line Israeli leaders, has created dry tinder that could burst into a major Palestinian uprising.
	TAPs transcript	Translator 1: ‘因为联系常识我们可以知道。巴勒斯坦和以色列之间确实是一直都有争议的问题，所以它大概的背景应该就是这件事情。’ (‘Given common knowledge, we can understand that there have indeed always been contentious issues between Palestine and Israel, so the general backdrop is likely this matter.’)
CASE 8	ST	Israel, which sees the weapons programme as an existential threat, has been <u>locked</u> for years in a shadow war with Iran.
	TAPs transcript	Translator 1: ‘封锁多年？我在想他是被封锁还是？被他国封锁？这里可能要需要一点的文化背景吧？’ (‘Blocked for years? I wonder whether it’s about being blocked or being blocked by another country? This might require some cultural background, right?’)

4.3. Translation Quality Analysis Across HT, NMT AND LLMs

A centrally positioned divergent bar chart is employed to represent Likert scale data, with the 0% mark serving as a neutral midpoint that indicates an equal balance of positive and negative responses. Bars extending to the right of the 0% point reflect a predominance of positive ratings towards the upper end of the Likert scale (scores 4 and 5), while bars stretching to the left indicate a predominance of negative ratings (scores 1 and 2). The length of each coloured segment within a bar indicates the

proportion of assessors that selected a particular rating (1 to 5). Figure 3 and Figure 4 depict the evaluation data across four distinct dimensions: accuracy, idiomaticity, vividness, and acceptability.

Figure 3 reveals that human translators achieve the highest mean scores across all four metrics, predominantly attributable to a distribution concentrated in the higher ends, specifically score 5. In contrast, the LLM and NMT groups exhibit lower average scores compared to human translators, and the data suggest that the LLM group demonstrates marginally superior

performance to NMT. Figure 4 illustrates the translation performance of six distinct subgroups. When comparing the performance of MAG and BAG, it is evident that MAG outperforms BAG on four dimensions, indicating that a longer period of time in translation training could enhance the improvement of translation quality. Intriguingly, the performance metrics of Youdao, GPT-3.5, and GPT-4 are on par with BAG in terms of accuracy, idiomatic expression, and vividness, which suggests that these

language models exhibit proficiency similar to that of novice translators. However, in terms of acceptability – which is inherently subjective and contingent upon the assessors' perceptions – it remains challenging for language models to achieve a level of performance comparable to human translators. Furthermore, DeepL demonstrates suboptimal performance in the translation task of MRWs, potentially due to its insufficient training on the English-Chinese pair.

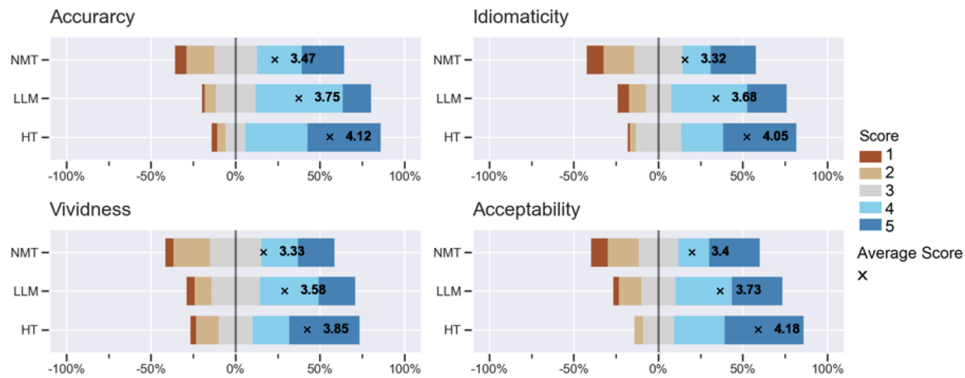


Figure 3. Likert scale plots for evaluation results across four metrics based on three groups HT, NMT and GPT

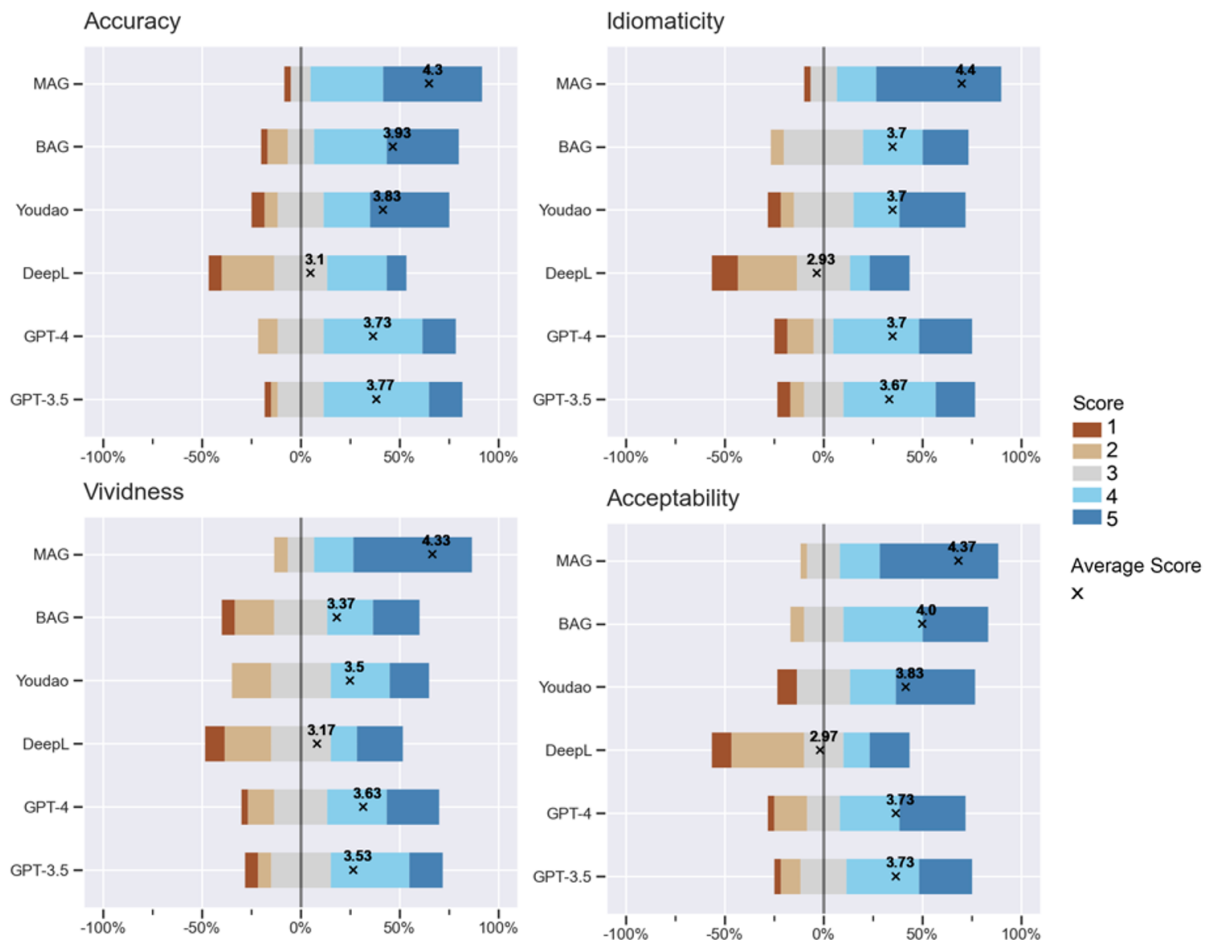


Figure 4. Likert scale plots for evaluation results across four metrics based on six subgroups MAG, BAG, Youdao, DeepL, GPT-4 and GPT-3.5

The distinct groups exhibit divergent patterns in the MRW translation task, and the evaluation data show variances in translation performance across these groups, as discussed in Section

4.1.1. It is plausible to assume that the choice of translation strategies has a significant impact on the assessment of translation, as explained in the subsequent analysis.

Table 7

*Comparative evaluation of translation strategies based on accuracy, idiomaticity, vividness, acceptability, and overall mean score*

STRATEGY	ACCURACY	IDIOMATICITY	VIVIDNESS	ACCEPTABILITY	MEAN
M→M (N = 49)	3.347	3.082	3.184	3.061	3.168
M→M2 (N = 41)	3.976	4.024	4.049	4.195	4.061
M→Non (N = 85)	3.965	3.882	3.671	3.988	3.876
M→∅ (N = 5)	3.200	3.400	2.400	3.600	3.150

Table 7 above demonstrates the average scores in four evaluation dimensions and overall average scores categorised by different translation strategies used. The overall mean values of the four dimensions provide an overview of the comprehensive performance among each translation strategy, with M→M2 (4.061) outperforming the others, followed by M→Non (3.876), M→M, and M→∅ (3.150). The average scores within the four

dimensions adhere to the similar trend. Table 5 presents the results of the Mann-Whitney U tests conducted across all six pairwise comparisons. It reveals that there are statistically significant divergences between the strategies M→M2 and M→M ( $p < 0.001$ ), as well as M→Non and M→M ( $p < 0.001$ ). This indicates that the choice of translation strategy exerts a discernible impact on the quality of the translation.

Table 8

*Statistical significance of differences between translation strategy pairs across various evaluation metrics*

PAIR	ACCURACY	IDIOMATICITY	VIVIDNESS	ACCEPTABILITY	MEAN
M→M vs M→M2	0.005**	0.001**	0.001**	0.000***	0.000***
M→M vs M→Non	0.001**	0.001**	0.024	0.000***	0.000***
M→M vs M→∅	0.828	0.592	0.175	0.319	0.964
M→M2 vs M→Non	0.693	0.377	0.058	0.220	0.219
M→M2 vs M→∅	0.113	0.102	0.008**	0.111	0.044*
M→Non vs M→∅	0.113	0.172	0.026*	0.308	0.076

\* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , calculated using the Mann-Whitney U test

Drawing upon the data presented in Tables 7 and 8, it can be concluded that the strategies of translating metaphors into alternative metaphors (M→M2) or into non-metaphorical expressions (M→Non) are favoured based on the evaluation from the perspective of assessors. This can be discerned from the higher mean scores and the statistically significant differences indicated by the Mann-Whitney U test results. The inference drawn from the results indicates that the HT group exhibits a higher propensity to employ the M→M2 strategy compared to the

other groups, which can be considered a contributing factor to the higher evaluation scores. Concurrently, the LLM group demonstrates a preference for the M→Non strategy, positioning it intermediate in the evaluation among the three groups. The NMT group, exhibiting a preference for the M→M strategy, correlates with lower scores, thus placing it at the lowest ranking. These observations provide insights into the relative effectiveness of different translation strategies employed by each group and their impact on the overall translation quality.

## 5. DISCUSSION

### 5.1. How strategies of translating metaphor vary from human translator to NMT and LLM

The product-oriented analyses reveal that in translating metaphors, human translators, NMT, and LLM show different translation patterns in choosing strategies. M→Non is the most frequently used strategy among human translators, both in BAG and MAG, and M→M2 is the second. The result aligns with the findings of a study by Pedersen (2017), which investigated how metaphors in 'Yes, Prime Minister' were subtitled from English into Swedish. Pedersen's research reveals that substitution (M→M2) and reduction (M→Non) are the primary strategies used to translate monocultural metaphors, with proportions of 32.7% and 30.3%, respectively. Viewing the human translator as a baseline, LLMs exhibit translation patterns more similar to human translators. In contrast, NMT systems tend to use the M→M strategy. A study further supports this, showing that Google Translate, an NMT system, prefers using 'reproduced metaphors' (M→M) in target texts (66.0%), consistent with our findings (Zajdel, 2022). The pattern varies in Youdao and DeepL, which may be attributed to differences in the corpora and training methods used. Although few studies have quantitatively analysed the translation strategies these systems use, especially in comparison to human translators, some studies show it is common for different NMT systems to show lexical, syntactic, and stylistic differences (Gao et al., 2024; Webster et al., 2020). Furthermore, Hendy et al. (2023) suggest that GPT translations, in comparison to NMT, process a nature of greater paraphrasticity. Ichien et al. (2023) assessed the ability of GPT-4 to provide natural language interpretations of novel literary metaphors, finding that human assessors rated metaphor interpretations generated by GPT-4 as superior to those provided by college students. Thus, it is confirmed that LLMs hold a paraphrastic tendency in metaphor translation tasks, and this tendency enables them to achieve good performance. This may explain why LLMs show a similar pattern in metaphor translation to that of human translators. However, as demonstrated in the case study, the performance of the two LLMs is unstable, with occasional failures to recognise metaphors.

In conclusion, LLMs exhibit similarities to human translators in choosing the strategy of translating metaphors, mainly the M→Non strategy, whereas NMT systems are more likely to use the M→M strategy. Hanks (2006) suggests that metaphoricity tends to increase when there is a lower degree of shared semantic properties between two concepts. Although the current study does not specifically explore the relationship between metaphoricity and translation strategies, this insight provides useful context for understanding metaphor translation.

### 5.2. What factors interfere with processing metaphors in translation?

As demonstrated in the process-oriented analyses, human translators process metaphor on a conceptual level, incorporating contextual factors and cultural awareness into the process of

metaphor comprehension and translation decisions. Although the experiment was not specifically designed to compare the translation processes of metaphorical and non-metaphorical language, the TAPs and interview transcripts show that some cognitive schemes interfere with understanding metaphors, for example, by triggering mental imaging and analogy. Other researchers have also identified distinct brain activation patterns associated with metaphorical language compared to non-metaphorical language, especially in processing novel metaphors, and it is evident that the process of metaphor comprehension engages a network of brain regions, drawing on a diverse array of cognitive resources distributed throughout the brain (Bambini et al., 2011). In addition to contributing to understanding metaphors, contextual factors and cultural awareness help translators make translation decisions. For example, in the interview transcript in Case 6, the translator mentions that the M→M strategy is used since the metaphor is universally understandable by any culture. This illustrates, from another perspective, that cultural disparities do not determine whether a metaphor is translatable but rather influence the approach to translating the metaphor. In the investigation of metaphor translation in *The Picture of Dorian Gray* from English to Spanish, Zajdel (2022) demonstrates that transcultural and monocultural metaphors show different preferences in the strategies used. Understanding the factors that interfere with metaphor processing in translation can provide valuable data to improve machine translation systems. For example, He et al. (2024) propose a framework that mimics the human translation process (encompassing knowledge mining, integration, and selection) to enhance LLMs translation quality. The results demonstrate improved performance under automated evaluations and indicate that this approach provides more favourable translations under human assessments.

### 5.3. Where do human translators, NMT and LLM meet/conflate against metaphor translation assessment metrics?

The assessment analyses indicate that, even though the quality of NMT systems and LLMs does not substantially exceed that of human translation, these language models possess a capacity for metaphor translation, from conventional to novel metaphors, which is commensurate with that of human translators. Specifically, Youdao, GPT-4, and GPT-3.5 reached comparable performances in metaphor translation compared to beginner-level human translators. However, the performance of DeepL does not achieve parity with that of the others, mostly because 70% of its target texts utilise M→M, which is considered to receive low scores as a translation strategy. This is contrasted with Hu and Li (2023) who find that DeepL performs well in translating literary texts from English to Chinese with an accuracy and fluency rate of above 80% compared to human translation as a baseline.

The analyses presented in the preceding sections suggest that human translators, NMT systems, and LLMs exhibit distinct translation patterns in processing metaphors. It is also observed that the choice of translation strategies considerably influences

the translation quality. This observation leads us to an inductive reasoning that the divergent translation performance among these entities may be fundamentally attributed to their respective translation patterns when dealing with metaphors.

## 6. CONCLUSION

With a focus on the translation of MRWs, this study employs a variety of research methodologies, including product analysis, think-aloud protocols, subsequent interviews, and translation quality evaluation, to uncover the process and product of human translation, NMT systems, and LLMs. The aim is to discern the variances in translation strategies and quality performances between humans and different AI systems through human ratings. The study shows that NMT systems and LLMs are capable of translating metaphors to a level comparable to human translators. This finding underscores the continued complexity of metaphor translation and highlights the nuanced challenges faces by machine translation systems.

## References

- Aghazadeh, E., Fayyaz, M., & Yaghoobzadeh, Y. (2022). Metaphors in pre-trained language models: Probing and generalization across datasets and languages. In S. Muresan, P. Nakov, & A. Villavicencio (Eds.), *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics* (pp. 2037-2050). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2022.acl-long.144>
- Bambini, V., Gentili, C., Ricciardi, E., Bertinetto, P. M., & Pietrini, P. (2011). Decomposing metaphor processing at the cognitive and neural level through functional magnetic resonance imaging. *Brain Research Bulletin*, 86(3-4), 203-216. <https://doi.org/10.1016/j.brainresbull.2011.07.015>
- Banerjee, S., & Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In J. Goldstein, A. Lavie, C.-Y. Lin, & C. Voss (Eds.), *Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for Machine Translation and/or Summarization* (pp. 65-72). Association for Computational Linguistics. <https://aclanthology.org/W05-0909>
- Briva-Iglesias, V., O'Brien, S., & Cowan, B. R. (2023). The impact of traditional and interactive post-editing on machine translation user experience, quality, and productivity. *Translation, Cognition & Behavior*, 6(1), 60-86. <https://dx.doi.org/10.1075/tcb.00077.bri>
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Cai, Z. G., Duan, X., Haslett, D. A., Wang, S., & Pickering, M. J. (2024). Do large language models resemble humans in language use? *arXiv*, Article 2303.08014. <https://dx.doi.org/10.48550/arXiv.2303.08014>
- Choi, M., Lee, S., Choi, E., Park, H., Lee, J., Lee, D., & Lee, J. (2021). MelBERT: Metaphor detection via contextualized late interaction using metaphorical identification theories. In K. Toutanova et al. (Eds.), *Proceedings of the 2021 Conference of the*
- North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1763-1773). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.naacl-main.141>
- Clark, E., Celikyilmaz, A., & Smith, N. A. (2019). Sentence mover's similarity: Automatic evaluation for multi-sentence texts. In A. Korhonen, D. Traum, & L. Márquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2748-2760). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1264>
- Croft, W., & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511803864>
- Dagut, M. (1987). More about the translatability of metaphor. *Babel*, 33(2), 77-83. <https://doi.org/10.1075/babel.33.2.06dag>
- Demjén, Z., & Semino, E. (2016). Introduction: Metaphor and language. In E. Semino & Z. Demjén (Eds.), *The Routledge handbook of metaphor and language* (pp. 1-10). Routledge.
- Do Dinh, E.-L., & Gurevych, I. (2016). Token-level metaphor detection using neural networks. In B. Beigman Klebanov, E. Shutova, & P. Lichtenstein (Eds.), *Proceedings of the Fourth Workshop on Metaphor in NLP* (pp. 28-33). Association for Computational Linguistics. <https://doi.org/10.18653/v1/W16-1104>
- Dove, G., & Fayard, A.-L. (2020). Monsters, metaphors, and machine learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-17). Association for Computing Machinery.
- Gao, R., Lin, Y., Zhao, N., & Cai, Z. G. (2024). Machine translation of Chinese classical poetry: A comparison among ChatGPT, Google Translate, and DeepL Translator. *Humanities and Social Sciences Communications*, 11(1), Article 835. <https://doi.org/10.1057/s41599-024-03363-0>
- Ge, M., Mao, R., & Cambria, E. (2023). A survey on computational metaphor processing techniques: From identification, interpretation, generation to application. *Artificial Intelligence Review*, 56, 1829-1895. <https://dx.doi.org/10.1007/s10462-023-10564-7>

- Ghazala, H. (2003). Idiomaticity between evasion and invasion in translation: Stylistic, aesthetic and connotative considerations. *Babel*, 49(3), 203-228. <https://dx.doi.org/10.1075/babel.49.3.03gha>
- Hale, S. B., & Campbell, S. (2002). The interaction between text difficulty and translation accuracy. *Babel*, 48(1), 14-33. <https://doi.org/10.1075/babel.48.1.02hal>
- Hanks, P. (2006). Metaphoricity is gradable. In A. Stefanowitsch & S. T. Gries (Eds.), *Corpus-based approaches to metaphor and metonymy* (pp. 17-35). Mouton de Gruyter. <https://dx.doi.org/10.1515/9783110199895.17>
- He, Z., Liang, T., Jiao, W., Zhang, Z., Yang, Y., Wang, R., ... & Wang, X. (2024). Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, 12, 229-246. [https://dx.doi.org/10.1162/tacl\\_a\\_00642](https://dx.doi.org/10.1162/tacl_a_00642)
- Hendy, A., Abdelrehim, M., Sharaf, A., Raunak, V., Gabr, M., Matsushita, H., ... & Awadalla, H. H. (2023). How good are GPT models at machine translation? A comprehensive evaluation. *arXiv*, Article 2302.09210. <https://dx.doi.org/10.48550/arXiv.2302.09210>
- Hu, K., & Li, X. (2023). The creativity and limitations of AI neural machine translation: A corpus-based study of DeepL's English-to-Chinese translation of Shakespeare's plays. *Babel*, 69(4), 546-563. <https://doi.org/10.1075/babel.00331.hu>
- Ichien, N., Stamenković, D., & Holyoak, K. J. (2023). Large language model displays emergent ability to interpret novel literary metaphors. *arXiv*, Article 2308.01497. <https://dx.doi.org/10.48550/arXiv.2308.01497>
- Jang, H., Jo, Y., Shen, Q., Miller, M., Moon, S., & Rosé, C. (2016). Metaphor detection with topic transition, emotion and cognition in context. In K. Erk & N. A. Smith (Eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics* (pp. 216-225). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P16-1021>
- Jankowiak, K., & Lehka-Paul, O. (2022). Novel metaphor translation is modulated by translation direction. *Applied Psycholinguistics*, 43(1), 177-192. <https://doi.org/10/gnp9n9>
- Jiao, W., Wang, W., Huang, J., Wang, X., Shi, S., & Tu, Z. (2023). Is ChatGPT a good translator? Yes with GPT-4 as the engine. *arXiv*, Article 2301.08745. <https://dx.doi.org/10.48550/arXiv.2301.08745>
- Junczys-Dowmunt, M., Dwojak, T., & Hoang, H. (2016). Is neural machine translation ready for deployment? A case study on 30 translation directions. *arXiv*, Article 1610.01108v3. <https://doi.org/10.48550/arXiv.1610.01108>
- Kloepfer, R. (1981). Intra- and intercultural translation. *Poetics Today*, 2(4), 29-37. <https://doi.org/10.2307/1772483>
- Kocmi, T., & Federmann, C. (2023). Large language models are state-of-the-art evaluators of translation quality. In M. Nurminen et al. (Eds.), *Proceedings of the 24th Annual Conference of the European Association for Machine Translation* (pp. 193-203). European Association for Machine Translation. <https://aclanthology.org/2023.eamt-1.19>
- Koglin, A. (2015). An empirical investigation of cognitive effort required to post-edit machine translated metaphors compared to the translation of metaphors. *The International Journal of Translation and Interpreting Research*, 7(1), 126-141. <https://trans-int.org/index.php/transint/article/view/369>
- Koglin, A., & Cunha, R. (2019). Investigating the post-editing effort associated with machine-translated metaphors: A process-driven analysis. *The Journal of Specialised Translation*, 31, 38-59.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Littlemore, J., & Fielden-Burns, L. V. (2023). On the fringes of metaphor: Using ambiguously figurative vague language to pragmatically negotiate sensitive topics in the English as a Medium of Instruction classroom. *Journal of Pragmatics*, 209, 1-14. <https://doi.org/10.1016/j.pragma.2023.02.016>
- Littlemore, J. (2019). *Metaphors in the mind: Sources of variation in embodied metaphor*. Cambridge University Press.
- Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Proceedings of the ACL-2004 Text Summarization Branches Out Workshop* (pp. 74-81). Association for Computational Linguistics. <https://aclanthology.org/W04-1013>
- Lin, Y. (2023). The relationship between machine translation and human translation in the era of artificial intelligence machine translation. *Applied and Computational Engineering*, 5(1), 133-138. <https://doi.org/10.54254/2755-2721/5/20230547>
- Lin, Z., Ma, Q., Yan, J., & Chen, J. (2021). CATE: A contrastive pre-trained model for metaphor detection with semi-supervised learning. In M.-F. Moens, X. Huang, L. Specia, & S. W. Yih (Eds.), *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (pp. 3888-3898). Association for Computational Linguistics. <https://dx.doi.org/10.18653/v1/2021.emnlp-main.316>
- Liu, Y., Iter, D., Xu, Y., Wang, S., Xu, R., & Zhu, C. (2023). G-Eval: NLG evaluation using Gpt-4 with better human alignment. In H. Bouamor, J. Pino, & K. Bali (Eds.), *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing* (pp. 2511-2522). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2023.emnlp-main.153>
- Mao, R., Lin, C., & Guerin, F. (2018). Word embedding and wordnet based metaphor identification and interpretation. In I. Gurevych & Y. Miyao (Eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics* (pp. 1222-1231). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P18-1113>
- Mao, R., Lin, C., & Guerin, F. (2019). End-to-end sequential metaphor identification inspired by linguistic theories. In A. Korhonen, D. Traum, & L. Márquez (Eds.), *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3888-3898). Association for Computational Linguistics. <https://doi.org/10.18653/v1/P19-1378>
- Mason, K. (1982). Metaphor and translation. *Babel*, 28(3), 140-149. <https://doi.org/10.1075/babel.28.3.05mas>
- Massey, G. (2021). Re-framing conceptual metaphor translation research in the age of neural machine translation: Investigating translators' added value with products and processes. *Training, Language and Culture*, 5(1), 37-56. <https://dx.doi.org/10.22363/2521-442X-2021-5-1-37-56>
- Massey, G., & Ehrensberger-Dow, M. (2017). Translating conceptual metaphor: The processes of managing interlingual asymmetry. *Research in Language*, 15(2), 173-189. <https://doi.org/10.1515/rela-2017-0011>
- Nacey, S., & Skogmo, S. F. (2021). Learner translation of metaphor: Smooth sailing? *Metaphor and the Social World*, 11(2), 212-234. <https://doi.org/10.1075/msw.00016.nac>

- Neidlein, A., Wiesenbach, P., & Markert, K. (2020). An analysis of language models for metaphor recognition. In D. Scott, N. Bel, & C. Zong (Eds.), *Proceedings of the 28th International Conference on Computational Linguistics* (pp. 3722-3736). International Committee on Computational Linguistics. <https://doi.org/10.18653/v1/2020.coling-main.332>
- Newmark, P. (1981). *Approaches to translation*. Pergamon Press Publishing.
- Nida, E. A. (1964). *Toward a science of translating: With special reference to principles and procedures involved in bible translating*. E.J. Brill. [http://traduttologiageneralenz.pbworks.com/w/file/attach/139290543/Nida--Toward\\_a\\_Science\\_of\\_Translating.pdf](http://traduttologiageneralenz.pbworks.com/w/file/attach/139290543/Nida--Toward_a_Science_of_Translating.pdf)
- Papineni, K., Roukos, S., Ward, T., & Zhu, W.-J. (2002). Bleu: A method for automatic evaluation of machine translation. In P. Isabelle, E. Charniak, & D. Lin (Eds.), *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics* (pp. 311-318). Association for Computational Linguistics. <https://doi.org/10.3115/1073083.1073135>
- Pedersen, J. (2017). How metaphors are rendered in subtitles. *Target*, 29(3), 416-439. <https://doi.org/10.1075/target.16038.ped>
- Pragglejaz Group. (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol*, 22(1), 1-39. <https://doi.org/10.1080/10926480709336752>
- Rojo, A. (2015). Translation meets cognitive science: The imprint of translation on cognitive processing. *Multilingua*, 34(6), 721-746. <https://doi.org/10.1515/multi-2014-0066>
- Schäffner, C. (2004). Metaphor and translation: Some implications of a cognitive approach. *Journal of Pragmatics*, 36(7), 1253-1269. <https://doi.org/10/fsz2z5>
- Sellam, T., Das, D., & Parikh, A. (2020). BLEURT: Learning robust metrics for text generation. In D. Jurafsky, J. Chai, N. Schuster, & J. Tetreault (Eds.), *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7881-7892). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.acl-main.704>
- Shuttleworth, M. (2011). Translational behaviour at the frontiers of scientific knowledge: A multilingual investigation into popular science metaphor in translation. *The Translator*, 17(2), 301-323. <https://dx.doi.org/10.1080/13556509.2011.10799491>
- Sjørup, A. C. (2013). *Cognitive effort in metaphor translation: An eye-tracking and key-logging study* (Publication No. 18.2013) [Doctoral dissertation, Copenhagen Business School]. Copenhagen Business School Publications. <https://research.cbs.dk/en/publications/cognitive-effort-in-metaphor-translation-an-eye-tracking-and-key->
- Snell-Hornby, M. (1988). *Translation studies: An integrated approach*. John Benjamins. <https://doi.org/10.1075/z.38>
- Song, W., Zhou, S., Fu, R., Liu, T., & Liu, L. (2021). Verb metaphor detection via contextual relation learning. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 4240-4251). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.acl-long.327>
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., & Krennmayr, T. (2010). Metaphor in usage. *Cognitive Linguistics*, 21(4), 765-796. <https://doi.org/10.1515/cogl.2010.024>
- Stowe, K., Chakrabarty, T., Peng, N., Muresan, S., & Gurevych, I. (2021). Metaphor generation with conceptual mappings. In C. Zong, F. Xia, W. Li, & R. Navigli (Eds.), *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing* (pp. 6724-6736). Association for Computational Linguistics. <https://dx.doi.org/10.18653/v1/2021.acl-long.524>
- Tendahl, M., & Gibbs, R. W. (2008). Complementary perspectives on metaphor: Cognitive linguistics and relevance theory. *Journal of Pragmatics*, 40(11), 1823-1864. <https://doi.org/10.1016/j-pragma.2008.02.001>
- Thibodeau, P. H., Matlock, T., & Flusberg, S. J. (2019). The role of metaphor in communication and thought. *Language and Linguistics Compass*, 13(5), Article e12327. <https://doi.org/10.1111/lnc3.12327>
- Tong, X., Choenni, R., Lewis, M., & Shutova, E. (2024). Metaphor understanding challenge dataset for LLMs. *arXiv*, Article 2403.11810. <https://doi.org/10.48550/arXiv.2403.11810>
- Tong, X., Shutova, E., & Lewis, M. (2021). Recent advances in neural metaphor processing: A linguistic, cognitive and social perspective. In K. Toutanova et al. (Eds.), *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4673-4686). Association for Computational Linguistics. <https://dx.doi.org/10.18653/v1/2021.naacl-main.372>
- Toury, G. (1985). Aspects of translating into minority languages from the point of view of translation studies. *Multilingua*, 4(1), 3-10. <https://doi.org/10.1515/mult.1985.4.1.3>
- Toury, G. (2012). *Descriptive translation studies – and beyond*. John Benjamins. <https://doi.org/10.1075/btl.100>
- Tran, N., Tran, H., Nguyen, S., Nguyen, H., & Nguyen, T. (2019). Does BLEU score work for code migration? In *Proceedings of the 2019 IEEE/ACM 27th International Conference on Program Comprehension (ICPC)* (pp. 165-176). IEEE. <https://doi.org/10.1109/ICPC.2019.00034>
- van den Broeck, R. (1981). The limits of translatability exemplified by metaphor translation. *Poetics Today*, 2(4), 73-87. <https://doi.org/10/dg9s9g>
- Vinay, J.-P., & Darbelnet, J. (1995). *Comparative stylistics of French and English: A methodology for translation*. John Benjamins. <https://doi.org/10.1075/btl.11>
- Wachowiak, L., & Gromann, D. (2023). Does GPT-3 grasp metaphors? Identifying metaphor mappings with generative language models. In A. Rogers, J. Boyd-Graber, & N. Okazaki (Eds.), *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (pp. 1018-1032). Association for Computational Linguistics. <https://dx.doi.org/10.18653/v1/2023.acl-long.58>
- Webster, R., Fonteyne, M., Tezcan, A., Macken, L., & Daems, J. (2020). Gutenberg goes neural: Comparing features of Dutch human translations with raw neural machine translation outputs in a corpus of English literary classics. *Informatics*, 7(3), Article 3. <https://doi.org/10.3390/informatics7030032>
- Zajdel, A. (2022). Catching the meaning of words. In J. Luke Hadley, K. Taivalkoski-Shilov, C. S. C. Teixeira, & A. Toral (Eds.), *Using technologies for creative text translation* (pp. 116-138). Routledge. <https://doi.org/10.4324/9781003094159-6>

Zhang, S., & Liu, Y. (2022). Metaphor detection via linguistics enhanced Siamese network. In N. Calzolari et al. (Eds.), *Proceedings of the 29th International Conference on Computational Linguistics* (pp. 4149-4159). ICCL.

Zheng, L., Chiang, W.-L., Sheng, Y., Zhuang, S., Wu, Z., Zhuang, Y., ... & Stoica, I. (2023). Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. *arXiv*, Article 2306.05685. <https://doi.org/10.48550/arXiv.2306.05685>

#### ABOUT THE AUTHORS

**Zhengjian Li**

BA in Translation, Department of Language Science and Technology, Saarland University, Saarbrücken, Germany

Postal address: Universität des Saarlandes, Campus, 66123 Saarbrücken, Germany

Email: [zheli@lst.uni-saarland.de](mailto:zheli@lst.uni-saarland.de)

ORCID: 0009-0001-5385-9236

**Lang Chen**

PhD in English Language and Literature, Professor of Applied Linguistics, School of Interpreting and Translation Studies, Guangdong University of Foreign

Studies, Guangzhou, China

Postal address: 2 Baiyun Avenue, Baiyun District, Guangzhou 510420, China

Email: [clheather8803@sina.com](mailto:clheather8803@sina.com)

ORCID: 0009-0009-2531-5950