SCIENTIFIC Publishing Limited

Digital Technologies Research and Applications
http://ojs.ukscip.com/index.php/dtra

Article

A Comparative Study of Human and Machine Translation of Animal Metaphors in Mo Yan's *Frog*

Juechu Yin [1] **and Qiushi Gu** [2,*]

[1] School of Foreign Languages, Jiangsu University, Zhenjiang 212000, China

[2] School of Japanese Studies, Beijing International Studies University, Beijing 100204, China

[*] Correspondence: guqiushi@bisu.edu.cn

Received: 20 November 2025; **Revised:** 12 December 2025; **Accepted:** 16 January 2026; **Published:** 30 January 2026

Abstract: Metaphor translation plays a key role in cross-cultural communication. Among metaphor types, animal metaphors stand out for their rich cultural connotations and cognitive complexity, making them a valuable testing ground for translation strategies. Despite growing interest, existing research has yet to fully clarify the cultural adaptation mechanisms involved in rendering animal metaphor translation across languages. In particular, how different translation agents dynamically process these culturally loaded expressions remains underexplored, which complicates efforts to optimize human-machine collaboration. This study adopts conceptual metaphor theory and an integrated methodology combining qualitative and quantitative analysis with theoretical interpretation. Drawing on three English translations of Mo Yan's *Frog*—by Howard Goldblatt, ChatGPT-4.0, and ChatGLM—this study conducts a systematic comparison of how human and machine translators handle animal metaphors. The analysis shows that effective rendering requires more than literal transfer: it depends on activating culture-specific frames and maintaining evaluative stance, not merely preserving surface imagery. While recent advances in artificial intelligence yield relatively high rates of literal retention, machine translations tend to remain surface-bound when metaphors are culturally or politically charged. By contrast, the cultural awareness and interpretive craft evident in the human translation more consistently preserve metaphorical nuance and ideological force. This study offers new evidence for research on metaphor translation and provides practical guidance for improving human-machine collaborations in literary contexts—e.g., using machine outputs to secure surface mapping while human translators recalibrate cultural frames and stance.

Keywords: Metaphorical Translation; Human-Machine Comparison; *Frog*; Animal Metaphors

1. Introduction

Metaphor is central to literary expression and remains difficult to carry across cultures. Mo Yan (Nobel Prize in Literature, 2012)'s *Frog* offers a dense field of animal metaphors whose cultural and ideological resonances make it well-suited for comparative analysis. In translation studies, recent scholarship on LLM (large language model)-mediated literal translation reports a tendency to keep surface imagery while under-realizing culture-specific entailments and ideological stance [1,2]. Given their dense cultural connotations and cognitive complexity, animal metaphors constitute a stringent stress test for machine translation [3–5]. To theorize these patterns, this study draws on Conceptual Metaphor Theory (CMT) to model source-to-target mappings and their systematicity, and on Eco's semiotics of context-bound interpretants to capture how cultural codes, intertextual frames, and reader knowledge shape interpretation—especially for animal imagery [6,7] (see Sections 2.1–2.3). Importantly, we make the central claim empirically checkable: rather than treating "faithfulness" as an overall impression, we operationalize translation performance along three aligned dimensions—mapping completeness (whether the species-level image is realized), frame retention (whether culture-specific entailments/allusions are preserved), and stance maintenance (whether attitudinal/ideological loading is maintained).

Empirically, we compare Howard Goldblatt's published English translation with outputs from ChatGPT-4.0 and ChatGLM (versions and settings detailed in Section 3.2). Using 52 metaphor contexts extracted from the Chinese source text, we report instance-level proportions with confidence intervals, paired contrasts, and a structured error analysis that identifies the conditions under which LLMs preserve mapping but attenuate frames and stance. This study addresses three research questions: (1) How do human and LLM translations diverge in handling animal metaphors? (2) To what extent do LLMs preserve lexicalized imagery, and under what conditions do they under-realize culture-specific frames and stance? (3) What follows from these patterns for designing more effective human–AI collaboration in literary translation (e.g., annotation targets, prompt design, division of labor)?

2. Literature Review

2.1. From Rhetorical Ornament to Meaning-Making in Context

Classical rhetoric—above all Aristotle and Quintilian—primarily treats metaphor as ornament or substitution, rather than as a mode of cognition. The interactionist accounts of the twentieth century (Richards; Black) reconceive metaphor as inter-term interaction and category projection, moving beyond decorative or taxonomic views. Yet these approaches offer limited analytical purchase on the cognitive operations and culture-bound codings that matter for cross-cultural literary translation, particularly where animal imagery carries socially shared scripts and evaluative load [8,9]. This limitation motivates frameworks able to address both conceptual mappings and semiotic circulation.

2.2. Cognitive Metaphor Theory (CMT)

Since *Metaphors We Live By*, metaphor has been understood as a cognitive mechanism that shapes perception and reasoning [1]; under this view, translation becomes a form of creative problem-solving rather than mere transfer. Building on this line, Zheng argues that metaphor translation engages experiential cognition—reading, comprehension, interpretation, and expression—and is therefore an inherently aesthetic activity [2]. At the same time, CMT shows recognizable limits in literary analysis when metaphors function as culturally coded sign systems. Prior work notes that CMT foregrounds putative universals of source-to-target mapping while under-specifying culture-specific frames and intertextual cues that guide interpretation [5,8]. Semiotic perspectives—most notably Eco's account of metaphors embedded in networks of cultural codes and interpretants—address these dimensions directly [10,11]. This distinction is salient for *Frog*: items such as the "toad" or the "landlord's dog" are not exhausted by generic mappings (e.g., animal—trait) but require cultural-historical contextualization of social types, genre memory, and evaluative stance [12].

Accordingly, this study adopts CMT as the primary explanatory frame for mapping regularities, while drawing on semiotic analysis to capture culture-specific entailments and code activation in animal metaphors. This combined lens aligns with the operationalization used later—mapping completeness, frame retention, and evaluative stance—and with the research questions set out in Section 1.

2.3. Metaphor Translation and the Emerging Gap in Human–LLM Comparison

Metaphor translation research has produced extensive typologies (e.g., conventional vs. novel metaphors; grammatical metaphor; etc.) and numerous literary case studies [3,6,7,13,14]. However, two gaps remain visible in the work most relevant to the present study. First, many literary analyses treat translation implicitly, offering rich readings without an operationally explicit account of how interpretive decisions are evaluated across languages and audiences [15–19]. Second, empirical comparisons between human translators and AI systems often prioritize surface-level features (lexical retention, grammatical fluency, or structural similarity) while leaving culturally embedded framing and evaluative load underspecified or unmeasured [20–22].

To address these gaps, the present study offers an exploratory but methodologically explicit design: we extract 52 animal-metaphor contexts from *Frog*, compare a canonical human translation with two LLM outputs, and evaluate all outputs using a pre-defined three-dimensional rubric that separates (i) mapping explicitness, (ii) frame retention, and (iii) stance maintenance. By reporting confidence intervals, paired contrasts, and error types, we contribute evidence that connects close reading to replicable empirical claims, thereby clarifying what current LLMs can and cannot do in culturally dense literary translation.

3. Methodology

This study compares a canonical human translation with two LLM outputs on animal metaphors in *Frog*. CMT guides the identification of mapping structures, while semiotic considerations guide the coding of culturally bound frames and stance (see Section 2).

3.1. Corpus Construction

The authors read the Chinese source text line by line and identified animal metaphors with the MIPVU procedure (established steps: determine lexical units, assign contextual vs. basic meanings, compare, then flag metaphorical use) [23]. We excluded zoologically literal uses, fixed proper names without metaphorical force, and immediate duplicates without new discourse function. The final dataset contains 52 unique metaphor contexts spanning 34 animal categories; because some contexts evoke multiple animal referents, token counts can exceed instance counts, and we report both denominators where relevant.

Two trained coders applied MIPVU independently (Cohen's $\kappa = 0.82$); disagreements were resolved through discussion. Each instance was stored with its surrounding context and chapter reference in a structured CSV to ensure traceability and consistent context windows during analysis. This operationalization follows integrated literary-cognitive treatments of figurative language [9].

3.2. Systems and Generation

This study uses Howard Goldblatt's published English translation as the human baseline. For LLMs, this study queried ChatGPT-4.0 and ChatGLM-130B to bracket English-centric vs. Chinese-optimized profiles. All 52 source snippets were translated in December 2024 via official web interfaces using a single instruction that situates the novel and requests literary consistency (verbatim wording available on request; see Data Availability Statement). Each system produced two runs under identical conditions; the first run is primary, the second checks stability. No post-editing beyond minimal formatting.

3.3. Coding Scheme and Statistical Analysis

We code each instance along three dimensions aligned with Section 2:

1. Mapping completeness: species-specific image realized (full/partial/none).
2. Frame retention: culture-specific entailments/allusions preserved (retained/attenuated/lost)
3. Evaluative stance: attitudinal/ideological load maintained (negative/neutral/positive).

Rubric transparency and bias control. All translations (human and LLM) were anonymized and randomized before coding so coders could not see the source of an output during evaluation. For frame and stance coding, the rubric specifies decision cues (e.g., dialectal insult triggers, political register markers, threat predicates, evaluative adjectives, intensity markers, and idiom-specific entailments). Disagreements were adjudicated against the rubric definitions rather than perceived literary quality.

Inter-coder agreement. Two coders independently coded all 52 contexts for all outputs; agreement for each dimension is reported in Results (κ for mapping/frame/stance).

Descriptive statistics and paired contrasts. We report instance-level proportions with Wilson 95% confidence intervals. For paired categorical contrasts between a given LLM and the human baseline (same source instance), we use McNemar tests and report effect sizes as odds ratios (OR) or risk differences (RD) with confidence intervals. For ordinal aggregates (e.g., retained/attenuated/lost), we report bootstrap CIs.

Given the modest number of instances ($n = 52$), we emphasize estimation (confidence intervals and effect sizes) to characterize uncertainty, and we interpret token-level summaries as descriptive rather than as standalone evidence for frame or stance fidelity.

Error analysis. To avoid over-reliance on token-level retention, we also conduct a structured error analysis: all cases of frame loss or stance weakening are categorized by trigger type (e.g., dialectal idiom, political slur, homophonic wordplay, narrative texture), with frequencies and representative examples reported in Section 4.

3.4. Materials and Transparency

The anonymized CSV (source snippet, context, identifiers), alignment keys to the published translation, system outputs from both runs, and the coding rubric are archived in a project repository and are available upon reasonable request for academic use, subject to copyright constraints on the full text.

4. Translation Analysis

This section answers the three research questions by aligning all evidence with the three analytic dimensions introduced in Sections 1–3: mapping completeness, frame retention, and stance maintenance. We treat the 52 unique metaphor contexts as the primary unit of analysis (instance level), because frames and stance are properties of contextual meaning rather than of isolated animal words. Token counts (when multiple animal images occur in one context or when images repeat) are reported only as a supplementary index of surface imagery and are not used to infer frame or stance fidelity.

4.1. Overall Results: Mapping, Frame, and Stance (Instance Level)

4.1.1. Mapping Completeness

In *Frog*, animal imagery recurs both as a stylistic resource and as a carrier of metaphorical meaning. To provide a transparent baseline for surface-image mapping, we first report token-level retention of animal-image lexemes across translations. The corpus contains 52 metaphor contexts (instances) and 82 animal-image tokens once repetitions and multi-referent contexts are counted. As shown in **Table 1**, token-level retention rates are similar across systems. Goldblatt retains 68/82 tokens (82.9%; Wilson 95% CI: 73.4–89.5), ChatGPT-4.0 retains 71/82 (86.6%; 95% CI: 77.6–92.3), and ChatGLM retains 70/82 (85.4%; 95% CI: 76.1–91.4). Given the overlap in confidence intervals, we treat token-level differences as descriptive. Each LLM output includes one context-level semantic mistranslation.

Table 1. Token-level surface imagery retention (mapping only; N-tokens = 82).

Parameter	Goldbatt (Human)	ChatGPT-4.0	ChatGLM
Metaphor contexts (unique contexts)	52	52	52
Animal-image tokens (incl. repeats/multi-referent)	82	82	82
Animal-image tokens retained, n	68	71	70
Retention rate (tokens retained/82), %	82.9%	86.6%	85.4%
Semantic mistranslations (in context), n	0	1	1
Wilson 95% CI, %	73.4–89.5	77.6–92.3	76.1–91.4

Note: Token counts exceed instance counts because some contexts contain repeated animal imagery or multiple animal referents. Percentages use N_tokens = 82 as the denominator. This table summarizes surface imagery retention (mapping) only; it does not measure whether culture-specific frames or evaluative stances are preserved, which require instance-level coding and are addressed in subsequent sections. Confidence intervals are reported for transparency and to avoid over-interpreting small token-level differences.

Importantly, token-level retention indexes map only (i.e., whether an animal lexeme is preserved) and do not by themselves capture frame retention or stance maintenance; these dimensions are evaluated at the instance level in the analyses that follow.

4.1.2. Frame Retention

Despite comparable token-level mapping (**Table 1**), frame retention is evaluated at the instance level, where culturally bound entailments and allusions are properties of contextual meaning rather than isolated animal words. Accordingly, we report instance-level frame retention outcomes (retained/attenuated/lost) with Wilson 95% CIs, and paired contrasts versus the human baseline using McNemar tests together with effect sizes (RD/OR) and their confidence intervals (**Table 2**). This pattern indicates that preserving a species term is not equivalent to preserving the cultural interpretant network that makes the metaphor do social and ideological work in context.

Table 2. Translation comparisons of the animal metaphor "耍死狗".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
田桂花，别耍死狗了，打你算轻的，应该送你进班房，从今后家里有生孩子的，都去找万医生。(Mo Yan, 2009, p.111)	Tian Guihua, stop the phony act. You got off lightly with only being struck. You ought to be sent to prison! From now on, Doctor Wan is to be called when any woman is about to have a child. Tian Guihua, if you ever again show up to do what you do, I'll rip those dog fingers right off your hands.	Tian Guihua, stop playing dead. Beating you was light; you should be thrown into jail. From now on, if anyone in your family is giving birth, they'll go to Doctor Wan. If you dare to deliver another baby, we'll chop off your dog's paws.	Don't play dead, dog, Tian Guihua. Hitting you is getting off easy; you should be sent to jail. From now on, anyone in the family who has a baby will go to Doctor Wan. Tian Guihua, if you dare to deliver another baby, I'll chop off your dog's paws.

4.1.3. Stance Maintenance

Stance maintenance is likewise assessed at the instance level. We therefore report the proportions of stance outcomes (maintained/weakened/strengthened) with 95% CIs, and paired effect sizes versus the human baseline (RD/OR with CIs) in **Table 2**. Across the dataset, the dominant drift is stance weakening—often through register normalization, dilution of insults, or replacement of politically loaded compounds with generic descriptors. To substantiate these claims beyond surface token retention, Section 4 further provides a structured error analysis of stance shifts by trigger type with frequencies and representative cases.

4.1.4. Representative Close Readings

As shown in **Table 2**, the source line and its three translations for "耍死狗" are juxtaposed for reference.

In "别耍死狗," the animal image functions as a dialectical scolding for feigned helplessness in a coercive power scene. Although LLM outputs can retain the lexical animal cue (dog), they tend to either drop the insult trigger (thereby weakening the stance) or produce an unidiomatic calque that misframes the idiom as literal performance. Goldblatt omits the animal lexeme at the trigger point but preserves the illocutionary force ("phony act") and compensates locally by reintroducing dehumanizing canine imagery within the same turn. This case illustrates a recurrent pattern in our dataset: token-level mapping is a poor proxy for fidelity when frame activation and stance carry the meaning.

Table 3 collates the source line and the three translations of "地主的狗崽子," which the close reading below analyses.

Table 3. Translation comparisons of the metaphor "地主的狗崽子".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
她说，她接生的第一个孩子本应该是革命的后代，没想到却接生了一个地主的狗崽子。(Mo Yan, 2009, p.110)	Her first ought to have been the son or daughter of a revolutionary, not a landlord's mongrel.	"She said the first child she delivered was supposed to be the offspring of revolutionaries, but unexpectedly, she ended up delivering the pup of a landlord."	She said that the first child she delivered should have been a descendant of the revolution, but to her surprise, she ended up delivering a bastard child of a landlord.

"地主的狗崽子" animalizes a class enemy within revolutionary discourse, encoding ideological heredity and contempt. LLM renderings that retain pup preserve mapping but often soften the political insult through affectively benign connotations, weakening stance; renderings that replace the animal metaphor with generic abuse preserve negativity but break the frame-bearing animalization logic. Goldblatt's choice preserves a hostile register and maintains the dehumanizing function even if English connotations shift slightly. This case supports the quantitative finding that LLMs can preserve animal lexemes while under-realizing historically and ideologically specific frames.

4.1.5. Interim Summary of Overall Results

Taken together, the token-level overview in **Table 1** and the two close readings ("耍死狗"; "地主的狗崽子") point to a consistent answer to RQ1. The LLM outputs are slightly more literal at the surface, reliably retaining the species word, yet they recurrently under-specify the culturally coded frames and pejorative stance that make these metaphors operative in context. The human translation, while marginally less "faithful" by a token metric, more steadily reconstructs the social scripts and the ideological work performed by animal imagery in *Frog*. In short, token-level retention is a serviceable index of mapping, but a weak proxy for functional fidelity once meaning depends on frame activation and evaluative polarity.

The next subsection turns from "how much is kept" to "under what conditions loss is most likely," beginning with a coarse split between universal and culturally bound items (**Table 2**), and then considering narrative position and novelty. This shift of focus clarifies why surface literalism travels easily while culturally embedded force does not.

4.2. Conditions under Which Imagery Is Preserved but Frames or Stance Weaken

This subsection builds on the overview in Section 4.1 and shifts the focus from the extent of retention to the conditions under which surface imagery travels while cultural frames or stance attenuate. We first present a split by cultural binding (**Table 2**), then examine narrative position and novelty, and conclude with two close readings that ground the pattern. These steps speak to the issues framed in RQ2.

4.2.1. Cultural Binding: Token-Level Distribution

To gauge the role of cultural binding, the corpus's 82 animal-image tokens (derived from 52 instances) were divided into universal items—broadly shared associations such as lion, dog, horse—and culturally bound items whose meanings are anchored in Chinese idiom, dialect, or political discourse (e.g., toad; landlord's dog). **Table 4** summarizes this split and reports token-level retention (i.e., mapping) rather than full fidelity.

Table 4. Token-Level Retention by Cultural Binding (animal-image tokens, N = 82).

Category	No. of Tokens	Retention (Goldblatt)	Retention (ChatGPT-4.0)	Retention (ChatGLM)
Universal metaphors	54	83.33%	85.19%	88.89%
Culturally bound metaphors	28	60.71%	64.29%	64.29%

Note: Counts are animal-image tokens (including repeats); totals sum to 82. "Universal" covers broadly recognizable associations; "culturally bound" covers usages rooted in Chinese frames or idioms. Token-level retention reflects surface imagery (mapping) only; frame and evaluative stance are analyzed qualitatively below.

As **Table 4** indicates, all three translations keep species words more readily in universal tokens. In culturally bound tokens, retention drops across the board, and the LLM advantage narrows. Crucially, the LLMs' numerical edge is largely a function of literal preservation of the animal lexeme; in culturally bound contexts, this often coincides with semantic flattening, where the social script, intertextual cue, or pejorative charge remains under-activated. By contrast, the human translation, although less "faithful" by a token metric, more consistently reconstructs idiomatic force in context. The figures thus motivate a closer look at two further variables—narrative position and novelty—that modulate frame and stance.

4.2.2. Narrative Exposition Versus Dialogue

Differences between narrative exposition and quoted dialogue affect how stance is signaled and, therefore, how it survives translation. In exposition, evaluation is frequently textured rather than explicit: it is distributed across syntax (e.g., paratactic piling, coercive imperatives), cadence (abrupt clause rhythms), and collocation (abuse terms paired with penal vocabulary). Such a texture is vulnerable to neutralizing paraphrase [24]. LLM outputs, even when lexically accurate at the species level, tend to standardize sentence architecture and soften the pragmatic edge; the result is a register shift from reprimand or political invective to a more neutral narrative voice [25].

By contrast, dialogue concentrates stance in vocatives, insults, and threats that are pragmatically foregrounded. Because the cues are overt and locally dense, both human and machine translations find it easier to preserve the negative polarity; the human version nevertheless shows greater sensitivity to how dehumanizing imagery and penal lexicon combine to build coercion. In short, stance is more fragile in exposition, where evaluation is a property of texture rather than of isolated words, and more resilient in dialogue, where the signal is explicit [9].

4.2.3. Novel Versus Conventional Metaphors

Novel metaphors—those without entrenched bilingual mappings—pose predictable difficulties across all three dimensions. Distributional learning favors conventional pairings (e.g., dog with cowardice, lion with courage), which stabilizes mapping. But when an animal image is mobilized against such default associations (e.g., lion used satirically of a domineering apprentice), or when its force relies on local intertext or dialectal scripts, the models often revert to safe literality and bland modifiers. The human translation, by contrast, may adjust the image or collocation to keep the function—the social script and attitudinal polarity—even at the cost of changing the surface mapping. For conventional metaphors, the principal risk is stance drift rather than loss of mapping; for novel ones, mapping, frame, and stance all become unstable.

4.2.4. Close Readings of Frame and Stance Shift

Novel metaphors—those without entrenched bilingual mappings—pose predictable difficulties across all three dimensions. Distributional learning favours conventional pairings (e.g., dog with cowardice, lion with courage), which stabilizes mapping. But when an animal image is mobilized against such default associations (e.g., lion used satirically of a domineering apprentice), or when its force relies on local intertext or dialectal scripts, the models often revert to safe literality and bland modifiers. The human translation, by contrast, may adjust the image or collocation to keep the function—the social script and attitudinal polarity—even at the cost of changing the surface mapping. For conventional metaphors, the principal risk is stance drift rather than loss of mapping; for novel ones, mapping, frame, and stance all become unstable. As shown in **Table 5**, the source line and the three translations of "小狮子" are presented for comparison.

Table 5. Translation comparisons of the animal metaphor "小狮子".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
姑姑带过 7 个徒弟，其中一个外号小狮子的，头发蓬松，塌鼻方口，脸上有粉刺，是姑姑的崇拜者。姑姑让她去杀人，她立马就会持刀前往，根本不问青红皂白。(Mo Yan, 2009, p.110)	She had seven interns, one of whom she called "Little Lion", a young woman whose hair was never combed, who had a flat nose, a square mouth, and a face full of zits. She was so devoted to Gugu that if she'd been told to kill someone, she'd have picked up a knife and done it without asking why.	Aunt had trained seven apprentices, one of whom was nicknamed "Little Lion." She had fluffy hair, a flat nose, a square mouth, and pimples on her face. She was an admirer of Aunt. If Aunt asked her to kill someone, she would immediately grab a knife and go, without asking any questions or seeking explanations.	Auntie has trained seven apprentices, one of whom is nicknamed "Little Lion." With fluffy hair, a flat nose, and a square mouth, her face dotted with acne, she is a devoted admirer of Auntie. If Auntie asked her to kill someone, she would grab a knife and go without a moment's hesitation, without questioning right from wrong.

As **Table 5** makes clear, applying "lion" to a woman carries aggressive, unruly overtones and supports satire of zeal. English defaults, however, are noble/courageous. The human translation steers the connotations toward irony by anchoring the epithet in negative collocations and by tightening the cadence around unthinking compliance. The LLM versions retain the animal word but allow default positive associations to bleed in, producing stance attenuation without any overt error at the level of mapping. The case exemplifies how novel or context-specific uses of an otherwise "universal" animal image are prone to semantic reversal in English unless framed by surrounding cues [5,8].

As shown in **Table 6**, the source line and the three English renderings of the homophonic sequence "蛙—娃—娲" are presented for comparison.

Table 6. Translation comparisons of the homophonic metaphor chain "蛙—娃—娲".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
常言道，蛙声如鼓，但姑姑说，那天晚上的蛙声如哭，仿佛是成千上万的初生婴儿在哭。(Mo Yan, 2009, p.169)	The croaking of frogs is often described in terms of drumbeats. But that night it sounded to her like human cries, almost as if thousands of newborn infants were crying.	"As the saying goes, the croaking of frogs is like the sound of drums, but Aunt said that on that night, the frogs' croaking sounded like crying, as if thousands of newborn babies were weeping."	As the saying goes, the croaking of frogs is like the sound of drums, but Auntie said that the sound of the frogs that night was like weeping, as if thousands of newborn babies were crying.

The homophonic chain links frog sound, infant crying, and mythic creation, converting animal imagery into an acoustic conduit for reproductive trauma. All English renderings preserve surface meaning but break the phonetic hinge that carries cultural resonance, resulting in frame attenuation irrespective of lexical accuracy. This case represents the error type "sound-based cultural coding," where literal mapping is structurally insufficient, and compensation strategies (paratext or intraline reformulation) become necessary.

4.2.5. Summary of Conditions and Patterns

Three conditions recur when surface imagery travels, but cultural force does not. First, cultural binding is the strongest predictor of loss: species-level mapping is portable, whereas social scripts, dialectical abuse, and intertextual cues require activation beyond the lexeme (**Table 4**). Second, narrative exposition is more vulnerable than dialogue, since stance is carried by texture and rhythm that generic paraphrase tends to neutralize. Third, novel metaphors are fragile across all dimensions; in conventional ones, stance drift is the main risk. In sum, cultural binding, narrative position, and novelty jointly shape where surface imagery travels, yet cultural force attenuates; the next section considers how such patterns can be addressed in practice without sacrificing literary nuance.

4.3. Implications for Human–AI Collaboration in Literary Translation

Findings from Sections 4.1–4.2 suggest a pragmatic division of labor. Large language models are dependable at species-level mapping in low-risk contexts, whereas cultural frames and evaluative stance remain most secure under human oversight. To make this operational, this study adopts a triadic QA checklist—mapping/frame/stance—used to review each token during copyediting, so that surface fidelity is evaluated alongside cultural activation and attitudinal polarity.

4.3.1. Workflow Design

Stage 1: Machine-first drafting for low-risk tokens.

Items with low cultural load and stable cross-lingual associations (e.g., size, speed, caution) may be drafted by an LLM to secure baseline mapping and terminological consistency. Prompts should remain minimal and context-aware (short scene brief; no stylistic over-specification), to avoid overwriting local voice.

Stage 2: Human determination of frame and stance for culturally coded tokens.

Where idiom, dialectal abuse, political coinage, or intertextual allusion governs meaning, translation decisions should begin by establishing the social script, activating intertext, and fixing attitudinal polarity. If literal imagery risks misdirection, functional equivalence is preferred, with local compensation (reintroducing the animal code nearby) or selective gloss at first mention. This stage treats mapping as subordinate to function.

Stage 3: Copy-editing with the triadic QA checklist.

Each animal-image token is audited against three prompts:

1. Mapping: Is the species-level image preserved where appropriate—or deliberately replaced for functional reasons?
2. Frame: Are culture-specific entailments (idiom, dialect, political register, intertext) activated for the target reader with minimal apparatus?
3. Stance: Does the translation maintain the intended polarity (pejorative/ironic/admiring), including through syntax, cadence, and collocation?

Disagreements are resolved at the context span (sentence/turn), not the word, to prevent token-level literalism from masking functional loss.

4.3.2. Applied Micro-Examples

As shown in **Table 7**, the source line and the three English renderings of "大洋马" are presented for comparison.

Table 7. Translation comparisons of the metaphor "大洋马".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
王胆，那么小个人，陈鼻，<u>大洋马</u>似的。(Mo Yan, 2009, p.128)	Wang Dan is so little, and Chen Bi is <u>as big as a horse</u>.	Wang Dan, such a small person, while Chen Bi was like <u>a big draft horse</u>.	Wang Dan, such a small person, Chen Bi, <u>as big as a draft horse</u>.

As **Table 7** indicates, "大洋马" (da yang ma; literally, a "foreign horse") is an epithet that encodes not only magnitude but an agrarian script that links labor, fecundity, and bodily instrumentality. Goldblatt's "as big as a horse" preserves the species image (mapping) while compressing the reproductive-political frame to size. ChatGPT/ChatGLM opt for "draft horse," which incidentally cues instrumentality yet still underactivates the gendered/fertility nuance associated with a mare and the social coloring of 洋 ("foreign, large-framed"). Where the scene thematizes reproductive control, a light adjustment—e.g., "breeding mare"—or a brief first-mention gloss can restore the intended frame without stylistic burden.

As detailed in **Table 8**, the source line appears with three English translations of "牛鬼蛇神."

Table 8. Translation comparisons of the metaphor "牛鬼蛇神".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
把牛鬼蛇神 拉上台来，萧上唇一声令下，那些严阵以待的纠察队员们两人挟持一个，将那些牛鬼蛇神脚不点地拥到了台上。(Mo Yan, 2009, p.126)	Bring the ox-ghosts and snake-demons up on the stage! Xiao Shangchun commanded. The miscreants' feet never touched the ground as they were bundled up onto the stage.	They dragged the "monsters and demons" onto the stage. At Xiao Shangchun's command, the vigilant inspection team members, two holding each one, lifted the 'monsters and demons' off their feet and onto the stage.	Drag the monsters and ghosts onto the stage, at the command of Xiao Shangshen, the disciplined squad members, who were standing by in strict formation, each pair grasped one, and without allowing their feet to touch the ground, they pushed and pulled the monsters and ghosts onto the stage in a rush.

Based on **Table 8**, "牛鬼蛇神" (niu gui she shen; literally, "ox-ghosts and snake-spirits"), the composite slur fuses animality with demonization to dehumanize political enemies [26]. Goldblatt's "ox-ghosts and snake-demons" retains the animal base (mapping) but may drift toward a Gothic reading unless the Cultural Revolution code is cued. The LLM outputs—"monsters and demons/ghosts"—standardize the phrase and erase the animal component, flattening the historically specific frame and softening stance from ideological vilification to generic monstrosity. A minimally invasive fix is lexical retention plus a first-mention cue (e.g., "a Cultural-Revolution slur for class enemies"), after which the compound can circulate without further apparatus.

Table 9 presents the source line together with three English treatments of "猫叫般的哭声."

Table 9. Translation comparisons of the metaphor "猫叫般的哭声".

Source Text	Howard Goldblatt	ChatGPT	ChatGLM
陈鼻刚出生时没有呼吸，姑姑将他倒提起来，拍打他的前胸后背，终于使他发出了猫叫般的哭声。(Mo Yan, 2009, p.110)	Chen wasn't breathing when he emerged, so Gugu held him by his feet and smacked him on the back and chest until he produced a kitten-like cry.	When Chen Bi was born, he wasn't breathing. Aunt held him upside down and patted his chest and back, finally making him let out a cry that sounded like a cat's meow.	When Chen Bi was born, he wasn't breathing. Auntie picked him up by his feet, patting his chest and back, until finally, he let out a cry that sounded like a cat's meow.

In **Table 9**, Mo Yan's childbirth scene, the animal metaphor "猫叫般的哭声" (mao jiao ban de ku sheng) literally means "a cat-like cry". In English, diminutives such as "kitten-like" may soften a stance by importing cuteness, and the idiom "the cat's meow" is a false friend with positive connotations. To preserve the scene's punitive register and the biopolitical coding of a precarious birth, prefer a non-diminutive, non-idiomatic collocation that keeps the sonic edge, e.g., "a catlike wail", "a rasping, catlike cry", or "a mewling, catlike cry." Pair the noun with verbs that sustain severity (e.g., wail/rasp/mewl rather than meow), and let surrounding penal collocations ("held him by his feet… smacked…") carry the coercive frame. This adjustment avoids idiomatic drift while retaining mapping (cat) and stabilizing stance (anguish rather than endearment).

4.3.3. Practical Implementation and Pedagogical Use

In project workflows, the triadic checklist can be instantiated as a one-page rubric attached to proofs: columns for mapping/frame/stance, checkboxes for kept/adjusted/compensated, and a line noting evidence of activation (e.g., idiom cue, political label, intertext). The unit of audit should be the context span (sentence or turn) rather than the isolated word, with brief rationales recorded for non-literal choices; periodic spot-checks support annotator calibration.

In teaching and training, the same triad structures peer review: a first pass marks mapping, a second annotates frames, and a third tests stance by paraphrasing the implied attitude and checking cadence/collocation. The sequencing discourages overreliance on token-level fidelity and trains attention on textural carriers of evaluation (syntax, rhythm, collocation).

4.3.4. Scope and Limits of the Workflow

The workflow presupposes access to a concise scene brief and tolerance for minimal paratext at first mention. In genres that proscribe notes, or where irony hinges on sound patterning (as in the 蛙—娃—娲 chain), compensation may need to occur within the line, and the tolerance for non-literal solutions correspondingly increases. The approach is model-agnostic: systems differ in style and register, yet show similar strengths at mapping; the triad provides a common evaluative lens across such variation. Taken together, these constraints suggest that the framework is best treated as a lightweight guide—useful where cultural activation and attitudinal polarity govern meaning, and to be adapted when genre or paratext conventions limit what can be signaled explicitly.

5. Conclusions

Grounded in CMT and informed by semiotic accounts of cultural code activation, this study examined how animal metaphors in Mo Yan's *Frog* are translated by a canonical human translator and two LLM systems. Using an explicit three-dimensional evaluation—mapping completeness, frame retention, and stance maintenance—we find a consistent split: LLMs typically preserve surface imagery at the lexical level, yet more frequently attenuate culturally bound frames and weaken evaluative stance, especially in dialectal insults, politically loaded compounds, sound-based wordplay, and passages where evaluation is carried by narrative texture rather than single words. By contrast, the human translation more consistently reconstructs social scripts and ideological work, including through functional equivalence and local compensation when literal mapping would mislead.

These findings contribute to translation studies and emerging AI-translation scholarship in two ways. First, they operationalize an often-intuitive claim ("image retained, frame/stance weakened") into replicable, instance-level evidence with confidence intervals, paired contrasts, and error-type analysis. Second, they clarify what "human–AI collaboration" can mean in practice: rather than treating LLM outputs as globally good or bad, translators can use LLMs for low-risk mapping while directing human effort toward frame and stance vulnerabilities that cluster in predictable trigger types.

Limitations. This study is based on a single novel and a corpus of 52 metaphor contexts, which constrains external validity and genre generalizability. Although we mitigate subjectivity through a pre-defined rubric, anonymized coding, and inter-coder reliability reporting, frame and stance coding inevitably retain interpretive components that may vary with readers' cultural knowledge. Finally, the LLM comparison is limited to two systems tested under a specific time window and interface conditions; model updates and alternative prompting strategies could shift absolute performance, even if the identified error types remain informative. Future research should expand across authors, genres, and metaphor classes, incorporate reader-response measures of perceived stance and cultural activation, and test whether targeted prompting or annotation can reduce the specific error clusters identified here.

Author Contributions

Conceptualization, J.Y. and Q.G.; methodology, J.Y. and Q.G.; validation, J.Y. and Q.G.; formal analysis, J.Y.; investigation, Q.G.; resources, J.Y.; data curation, J.Y.; writing—original draft preparation, J.Y.; writing—review & editing, Q.G.; visualization, J.Y.; supervision, Q.G.; project administration, Q.G. Both authors have read and agreed to the published version of the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The original corpus is not publicly available due to privacy and ethical restrictions copyright, and licensing considerations. For further inquiries regarding the data, please contact the corresponding author.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Lakoff, G.; Johnson, M. *Metaphors We Live By*; University of Chicago Press: Chicago, IL, USA, 1980.
2. Zheng, H.L. The essential characteristics of literary translation. *Chin. Transl. J.* **1998**, *6*, 5–9.
3. Huang, G.W. The analysis of grammatical metaphor in translation studies. *Chin. Transl. J.* **2009**, *1*, 1–5.
4. Rouhi, M.; Mahand, M.R. Animal metaphor in cognitive linguistics. *Psychol. Res.* **2011**, *1*, 251–254.
5. Talebinejad, M.R.; Dastjerdi, H.V. A cross-cultural study of animal metaphors: When owls are not wise! *Metaphor Symbol* **2005**, *20*, 133–150. [CrossRef]
6. Xue, M. A study of translation of novel metaphor from Chinese into English. *Shanghai J. Transl.* **2019**, *1*, 25–30.
7. Chen, J. A cross-language study on vision based-metaphors in English and Chinese. *J. PLA Univ. Foreign Lang.* **2003**, *1*, 20–22, 27.
8. Goatly, A. Humans, animals, and metaphors. *Soc. Animals* **2006**, *14*, 15–37.
9. Hart, K.R.; Long Jr., J.H. Animal metaphors and metaphorizing animals: An integrated literary, cognitive, and evolutionary analysis of making and partaking of stories. *Evol. Educ. Outreach* **2011**, *4*, 52–63. [CrossRef]
10. Eco, U. *A Theory of Semiotics*; Indiana University Press: Bloomington, IN, USA, 1976.
11. Eco, U. *Experiences in Translation*; University of Toronto Press: Toronto, ON, Canada, 2001.
12. Zhang, H. Metaphorical mapping and translation of animal cultural connotations in Chinese and English in intercultural communication. *Zool. Anim. Physiol.* **2021**, *2*, 50–60. [CrossRef]
13. Chen, J.X. A contrastive study of metaphorical cognition in English and Chinese. PhD Thesis, East China Normal University, Shanghai, China, 2004.
14. Lan, C. A cognitive approach to spatial metaphors in Chinese. *Foreign Lang. Teach. Res.* **1999**, *4*, 7–15.
15. Liu, W.; Yu, Z. Metaphorical translation strategies of the white paper "China's Green Development in the New Era." *Forum Linguist. Stud.* **2023**, *5*, 1858. [CrossRef]
16. Cao, J.; Graham, C. Metaphorical conceptions of time: A corpus-based semantic analysis of "TIME IS MONEY" in English and Chinese. *Forum Linguist. Stud.* **2025**, *7*, 520–531. [CrossRef]
17. Muhajir; Sumarwati; Saddhono, K. Linguistic symbolism and conceptual metaphors in Javanese Islamic myths: A semiotic perspective. *Forum Linguist. Stud.* **2025**, *7*, 77–89. [CrossRef]
18. Baker, M. *Translation and Conflict: A Narrative Account*; Routledge: London, UK, 2006.
19. Algobaei, F.; Alzain, E. Metaphors in translation: A study of figurative language in Badriah Albeshr's Hend and the Soldiers. *Forum Linguist. Stud.* **2025**, *7*, 740–763. [CrossRef]
20. Zhu, Y.F.; Guan, X.C. Translation education in the AI era: Challenges and opportunities. *Contemp. Soc. Sci.* **2019**, *27*, 37–45.
21. Weng, Y.M.; Wang, J.P. Characteristics and Limitations of Machine Translation of Literary Texts: A Comparative Study of Human-Machine Translation of Chinese Flowing Sentences. *Contemp. Foreign Lang. Stud.* **2020**, *20*, 128–137. [CrossRef]
22. Li, F.Q. A comparative study on the quality of English-Chinese translation between translation learners and a machine translation system in the era of artificial intelligence. *Foreign Lang. World* **2022**, *4*, 72–79.
23. Pragglejaz Group. MIP: A method for Identifying Metaphorically Used Words in Discourse. *Metaphor Symbol* **2007**, *22*, 1–39. [CrossRef]
24. Foucault, M. *Discipline and Punish: The Birth of the Prison*; Gallimard: Paris, France, 1975.
25. Lefevere, A. *Translation, Rewriting, and the Manipulation of Literary Fame*; Routledge: London, UK, 1992.
26. Alwash, L.M.A.; Abdul Jabar, M.A.; NimehchiSalem, V.; et al. The conceptual metaphor of animals used for praise and satire in Al Ghaddafi's discourse. *Int. J. Acad. Res. Prog. Educ. Dev.* **2023**, *12*, 1532–1549. [CrossRef]