

Human-in-the-Loop Cultural Schema Annotation for Metaphor Translation: A Pilot Framework

DAHUI DONG

Foreign Language School
Guangzhou Institute of Science and Technology
Guangzhou City, Guangdong Province,
CHINA

ORCID: 0000-0002-0607-9077

Abstract: Metaphor translation presents a persistent challenge in cross-lingual communication because source and target languages often organize metaphorical meaning through different cultural schemas. This pilot study proposes a human-in-the-loop annotation framework for tracing cultural schema transfer in Chinese-English metaphor translation. The framework annotates 500 Chinese-English metaphor pairs across four dimensions: source-language cultural schema, target-language cultural schema, transfer strategy, and translator proficiency level. BERT-wwm-assisted pre-annotation is used to generate candidate labels, which trained human annotators revise and adjudicate. An exploratory CLIP-based similarity measure is used only as a heuristic indicator of schema alignment, and a Graph Neural Network models associations among schema labels, proficiency levels, and observed transfer-strategy labels. In the pilot corpus, the GNN reached 82% accuracy in predicting adjudicated transfer-strategy labels. Corpus patterns indicated that intermediate translators favored literal translation (68%) when schema similarity was low, whereas advanced translators more often used adaptation or substitution strategies (74%). These findings are presented as preliminary evidence for the feasibility of schema-aware annotation and label prediction, not as evidence of downstream translation-quality gains or broad model generalization. The study contributes a reproducible pilot framework for making cultural-schema decisions more explicit in translation pedagogy and computational annotation while identifying the validation steps required for larger-scale work.

Key-Words: Cultural Schema Transfer; Metaphor Translation; Human-in-the-Loop Annotation; Parallel Corpus Annotation; Chinese-English Translation; Translator Proficiency

Received: March 23, 2026. Revised: May 27, 2026. Accepted: June 28, 2026. Published: October 5, 2026.

1 Introduction

Metaphor translation remains a difficult problem in cross-linguistic communication between Chinese and English. Prior work on Transformer-based models reports high lexical accuracy, [1], but culturally embedded metaphors remain difficult, [2], especially when source and target schemas diverge, [3]. The difficulty stems from cultural schemas embedded in metaphorical expressions, which technical equivalence alone cannot resolve. In translator training, explicit schema awareness can help distinguish mechanically adequate renditions from culturally appropriate ones.

Cognitive linguistics shows that metaphors structure thought through conceptual mappings, [4], [5]. Cultural schema research complements this by revealing that conceptual mappings are culturally situated, not universal, for example in Confucian versus Western schema systems, [6], [7], [8]. Parallel corpus methodology has proven instrumental for studying metaphor transfer through annotation approaches, [9], [10], and documented schema preservation patterns, [11], [12]. Translation

pedagogy research demonstrates that intermediate learners default to literal translations while advanced learners employ adaptation or substitution strategies, [13], suggesting that schema awareness is learnable across proficiency levels, [14].

Despite these advances, three gaps remain. Existing annotation schemes are fragmented: some corpora focus on conceptual mappings, [15], others prioritize pragmatic strategies like domestication, [16], but few integrate cultural schemas with translator proficiency data, [14]. Translation frameworks often treat cultural context as peripheral; most linguistic models focus on surface-level equivalences without systematic schema analysis. Computational systems struggle with metaphor translation because parallel corpora rarely annotate schema layers explicitly, [10]. While NLP techniques handle metadata, [17], and image schemas, [18], neural models lack the cultural grounding necessary for culturally appropriate output.

This paper proposes a multi-layered annotation framework that addresses these gaps by unifying cultural schema analysis, transfer strategy identification, and proficiency-level annotation.

By treating these dimensions as interdependent rather than isolated, the framework enables analysis of how cultural schemas are associated with translation choices across proficiency levels. The technical contribution is deliberately framed as a pilot human-in-the-loop annotation pipeline rather than a fully autonomous translation system: BERT-wwm-assisted pre-annotation generates candidate schema labels, human annotators refine and adjudicate those labels, and a Graph Neural Network models schema-strategy-proficiency associations. In this sense, the study operationalizes selected aspects of cultural schema transfer as observable annotation categories and empirically modelable relationships.

2 Cultural Metaphor in Translation

The study of metaphor translation has evolved over decades, shifting from linguistic analysis to approaches incorporating cognitive and cultural dimensions. Early work treated metaphors as lexical phenomena, focusing on surface-level equivalences between languages, [19]. Traditional studies examined how Chinese proverbs could be rendered into English through literal or paraphrased translations, [20]. While these methods provided foundational insights, they overlooked the deeper cultural schemas that give metaphors their meaning.

2.1 The Cognitive Turn in Metaphor Studies

Cognitive linguistic work redefined metaphors as conceptual mappings rather than linguistic ornaments, [21], [5]. This perspective showed how metaphors structure human thought and are embedded in cultural cognition. For translation, rendering a metaphor accurately required understanding the underlying conceptual domains. For example, the Chinese metaphor “雪中送炭” (timely help in need) maps to “timely help,” but the English “a friend in need” operates within a different cultural frame. Cognitive approaches thus emphasized recognizing and bridging these conceptual gaps, [4].

However, cognitive models initially paid limited attention to the cultural specificity of conceptual mappings. While they acknowledged that metaphors vary across languages, they did not systematically account for how cultural schemas (shared knowledge structures about social norms, values, and practices) shape these variations. This gap became particularly evident in studies of Chinese-English translation, where metaphors often encode Confucian, Daoist, or other culturally specific schemas, [22].

2.2 The Emergence of Schema-Oriented Approaches

Recent work integrates cultural schemas into metaphor translation research, recognizing that conceptual mappings are culturally situated, not universal. Prior schema-theoretic research, [23], proposed a framework for analyzing how translators navigate culturally bound metaphors by identifying and adapting source-language schemas to target-language contexts. This aligns with longer-standing concerns in translation studies about cultural mediation, translator choice, and target-culture norms, [24], [25], [26], particularly in non-Western languages where metaphors draw on historical or philosophical references, [27].

Parallel corpus research has been instrumental in advancing schema-oriented approaches. By annotating metaphors with cultural and conceptual tags, prior work has identified how translators negotiate schema mismatches, [12]. For instance, a study of Chinese political discourse found that metaphors rooted in revolutionary rhetoric were often adapted or omitted in English translations to align with Western democratic schemas, [28]. These findings highlight the importance of schema awareness in producing culturally coherent translations.

Nevertheless, existing annotation schemes remain fragmented. Some corpora focus on conceptual mappings, [15], while others prioritize pragmatic strategies like domestication or foreignization, [16]. Few integrate both dimensions with translator proficiency data, despite evidence that schema transfer competence develops unevenly across learning stages, [14]. This fragmentation limits the utility of annotated corpora for both pedagogy and computational modeling.

2.3 Theoretical Foundations: Human-in-the-Loop Support in Metaphor Translation

Recent advances in human-AI research provide useful foundations for designing annotation workflows in metaphor translation. Coordinated human-AI systems can show synergistic effects in learning environments, [29]. Bayesian modeling has also shown that human expertise and machine predictions may be complementary, [30]. Interactive AI-assisted modes can outperform isolated AI-only approaches in some academic writing contexts, [31]. For metaphor translation, where cultural schemas require both human interpretive judgement and computational pattern recognition, these findings support a human-in-the-loop design in which the machine proposes candidate patterns and human

annotators retain interpretive authority.

This framing also requires conceptual caution. Recent language-education research, [32], distinguishes genuine human-AI collaboration from more limited human-AI interaction by identifying features such as shared goals, interdependence, reciprocal communication, distributed accountability, complementary expertise, calibrated trust, and acknowledged contribution. The present study does not claim that its pilot pipeline satisfies all of these conditions. Human involvement is centered on correction, adjudication, and validation of annotation labels, while the AI components support candidate generation and pattern modeling. The framework is therefore more accurately described as a human-in-the-loop annotation framework with collaborative aspirations, rather than a fully reciprocal collaborative system. This distinction is important for translator training because it keeps the pedagogical goal focused on making cultural-schema reasoning explicit without overstating the agency of the computational components.

2.4 The Proposed Approach

The proposed framework addresses these limitations by unifying cultural schema analysis with transfer strategy and proficiency-level annotations. Unlike prior work treating these dimensions in isolation, our approach explicitly models their interactions. For example, preliminary data suggest intermediate translators default to literal translations when facing culturally opaque metaphors, whereas advanced translators employ schema substitution or adaptation, aligning with earlier findings, [14].

The framework's novelty lies not in proposing cultural schemas as stable computational objects, but in making selected schema-related translation decisions explicit, comparable, and empirically traceable. By decomposing translation decisions into annotated layers, it provides a structured way to examine how schema labels, transfer strategies, and proficiency levels co-occur in a pilot corpus. This contrasts with conventional transfer-based approaches, [33], focusing on lexical or syntactic equivalences, cognitive models, [3], prioritizing conceptual mappings without schema tagging, and hybrid methods, [2], annotating strategies and schemas separately without considering interdependencies. The GNN-based component should therefore be understood as modeling observed schema-strategy-proficiency associations, not as proving that cultural schemas are universally stable or automatically transferable across cultures.

Implementing this framework requires addressing several technical and methodological challenges. Cultural schemas are context-dependent constructs,

so their annotation depends on transparent guidelines, inter-annotator calibration, and cautious interpretation. Pre-trained embeddings such as CLIP may inadvertently encode cultural biases, [34], and their use for abstract cultural semantics should be treated as exploratory. Proficiency-level classification necessarily simplifies translator competence, [35], which often varies across domains and metaphor types. The framework provides a methodological starting point for these issues, but broader validation will require larger corpora, additional language pairs, and independent translation-quality evaluation.

The original contribution of this study is therefore bounded but specific. It does not claim to introduce a new metaphor theory, a new CLIP architecture, or a fully validated translation-quality predictor. Its contribution is to define an integrated annotation and label-prediction workflow in which source schema, target schema, transfer strategy, and translator proficiency are represented jointly and checked through human adjudication. This makes the study relevant to computational and mathematical audiences as an operational modeling problem, while preserving its grounding in translation studies and cultural-schema theory.

3 Cultural Schemas and Parallel Corpus Annotation

Understanding metaphor translation requires more than linguistic proficiency; it demands awareness of the cultural schemas that shape metaphorical meaning. These schemas (cognitive frameworks that organize cultural knowledge) play a central role in how metaphors are interpreted and rendered across languages. When combined with parallel corpus annotation, they offer a structured way to analyze and teach translation strategies.

3.1 Cultural Schemas in Linguistics

Cultural schemas function as mental blueprints that guide how individuals perceive and articulate experiences within their cultural context. Rooted in anthropology and cognitive linguistics, the concept explains why certain metaphors resonate deeply in one culture but fall flat in another. For example, the Chinese metaphor “画蛇添足” (drawing legs on a snake) encodes a cultural schema about unnecessary excess, derived from Confucian teachings on moderation, [6]. In contrast, English speakers might use “gilding the lily” to convey a similar idea, reflecting a different aesthetic and moral framework. Research on cultural transfer mechanisms supports the need for human interpretive mediation: studies of crisis translation show that

informed human mediators can improve culturally appropriate outputs for culture-specific metaphors, [36]; cross-linguistic schema divergences require explicit annotation, [7]; and foundational work maps Confucian versus Western schema systems, [6], [8]. This evidence indicates that cultural schema transfer should not be treated as fully automatable. Human-in-the-loop pipelines can make cultural mediation more systematic while preserving the role of expert judgement in metaphor-rich texts.

Such schemas are not monolithic but vary across domains like social hierarchy, morality, and time perception. Studies in cross-cultural pragmatics show that translators often struggle when source and target cultures hold conflicting schemas for the same concept, [37]. For instance, translating Chinese familial metaphors into English requires navigating differing schemas of filial piety, where Chinese culture emphasizes collective responsibility while Western cultures prioritize individual autonomy, [8]. These mismatches highlight the need for schema-aware translation methods that go beyond lexical substitution.

3.2 Parallel Corpus in Translation Studies

Parallel corpora have become indispensable tools for studying how metaphors travel across languages. By aligning source texts with their translations, researchers can identify patterns in how cultural schemas are preserved, adapted, or omitted. Early corpus-based studies focused on lexical and syntactic equivalences, [38], but recent work has expanded to include cultural and conceptual annotations, [11]. For example, the Chinese-English Metaphor Corpus tags metaphors with both conceptual domains (e.g., “time as space”) and cultural schemas (e.g., “cyclical time in Daoism”), enabling fine-grained analysis of translation choices, [7].

However, constructing such corpora poses challenges. Annotators must balance granularity with consistency, ensuring that schema labels are neither overly broad nor idiosyncratic. Inter-annotator agreement is particularly difficult for culture-specific metaphors, where even native speakers may interpret schemas differently, [39]. Computational tools like word embeddings have aided schema identification, [40], but human expertise remains important for context-dependent cases, such as metaphors drawing on historical or literary allusions.

3.3 Pedagogical Foundations of Translation Training

Traditional translation pedagogy often treats cultural competence as an adjunct to linguistic training, despite evidence that schema awareness distinguishes proficient translators. Research in

translator education shows that intermediate learners tend to rely on literal translations when faced with schema mismatches, whereas advanced learners employ strategies like adaptation or explicitation, [13]. For example, an intermediate translator might render “red-eye disease” (red-eye disease, denoting envy) literally, while an advanced translator would opt for “green-eyed monster,” adapting the schema to the target culture’s idiomatic repertoire.

This progression aligns with schema theory’s emphasis on knowledge restructuring. As translators encounter more metaphors, they develop a “schema repertoire” that helps them predict and resolve cultural mismatches, [41]. Pedagogical interventions such as contrastive analysis of parallel texts or schema-mapping exercises can accelerate this development. For instance, trainees might compare how Chinese and English news metaphors frame economic growth, analyzing how collectivist versus individualist schemas influence lexical choices, [42].

Integrating schema awareness into curricula also addresses a critical gap in machine translation training. While neural models excel at surface-level equivalences, they often falter with schema-rich metaphors unless explicitly exposed to culturally annotated data, [18]. This suggests a practical connection between human pedagogical practice and computational model development: human translators can leverage schema knowledge to make context-sensitive decisions, while annotated corpora can systematize the kinds of schema distinctions that models otherwise miss. The present study focuses on building and testing such an annotation structure rather than evaluating downstream machine-translation improvement.

4 A Multi-layered Annotation Scheme for Chinese-English Metaphor Translation

The proposed annotation framework systematically addresses the challenges of cultural schema transfer in metaphor translation by integrating four dimensions: source and target cultural schemas, transfer strategies, and translator proficiency levels. This section details the technical implementation of each component and their interplay in the annotation process.

4.1 Four-Dimensional Cultural Schema Annotation

The annotation scheme operationalizes selected aspects of cultural schema transfer through explicit tagging of metaphors in parallel corpora. Let $D = \{(m_i, m'_i, x_i)\}_{i=1}^N$ denote the pilot corpus, where m_i is a Chinese source metaphor, m'_i is its English

rendering, x_i is the local textual context used during annotation, and $N = 500$ in the present study. Each pair is assigned four annotation dimensions by trained annotators after machine-assisted pre-annotation and human adjudication.

1. Source-Language Cultural Schema (SCS)

$$SCS(m_i) = \{Domain : d_{s,i}, Schema : s_{s,i}\} \quad (1)$$

where $d_{s,i}$ denotes the source-language cultural domain (e.g., ConfucianMorality) and $s_{s,i}$ the specific source schema (e.g., FilialPietyHierarchy).

2. Target-Language Cultural Schema (TCS)

$$TCS(m'_i) = \{Domain : d_{t,i}, Schema : s_{t,i}\} \quad (2)$$

where $d_{t,i}$ and $s_{t,i}$ are the target-language counterparts identified in the English rendering. For example, the Chinese “三纲五常” (Three Bonds and Five Constants) might map to $d_t = WesternIndividualism$ and $s_t = SocialContract$ when the translation reframes the source schema through a target-culture social-order concept.

3. Transfer Strategy (TS)

$$T(m_i, m'_i) \in \{Literal, Adaptation, Substitution, Omission\} \quad (3)$$

This captures how m_i is rendered into m'_i , with Adaptation indicating partial schema alignment (e.g., “画蛇添足” → “gilding the lily”) and Substitution denoting full schema replacement (e.g., “红眼病” → “green-eyed monster”).

4. Translator Proficiency Level (TPL)

$$TPL_i \in \{Intermediate, Advanced\} \quad (4)$$

Annotated based on the translator’s experience and corpus metadata, as proficiency influences strategy choice.

These dimensions are interdependent: TS is conditioned on SCS, TCS, and TPL, forming a probabilistic relationship:

$$P_\theta(T_i | SCS_i, TCS_i, TPL_i, x_i) \quad (5)$$

where $T_i = T(m_i, m'_i)$ and θ denotes the parameters of the label-prediction model.

Equation (5) is not intended as a causal model of translation behavior. It formalizes an associational prediction task: given adjudicated schema labels, local context, and proficiency metadata, the model estimates the probability of an observed transfer-strategy label. This distinction is important because the pilot corpus is cross-sectional and does not support causal inference about how proficiency changes translation behavior over time.

Table 1 (Appendix) summarizes the notation used in Eqs. (1)–(5) and in the computational components introduced below.

4.2 CLIP-Enhanced Schema Similarity

To provide an exploratory measure of schema alignment, we compute the similarity between $SCS(m)$ and $TCS(m')$ using CLIP embeddings. CLIP was not originally designed to model abstract cultural semantics; therefore, in this pilot study it is used only as a heuristic similarity feature, not as a definitive measure of cultural equivalence. Each schema label was converted into a short cultural description, and the source and target descriptions were embedded and compared. For a metaphor m with source-schema description c_m and target-schema description $c_{m'}$, the exploratory schema-similarity score is

$$Sim_{schema}(m, m') = \cos(CLIP(c_m), CLIP(c_{m'})) \quad (6)$$

For example, c_m for “雪中送炭” might be “help in dire need,” while $c_{m'}$ for “a friend in need” might be “mutual aid in hardship.” Higher scores were interpreted as tentative evidence of schema preservation, while lower scores suggested possible Substitution or Omission. Final interpretation of schema similarity remained subject to human adjudication, because the embedding score may reflect cultural and linguistic biases in the underlying model.

4.3 GNN-Based Strategy Prediction

The annotated data train a Graph Neural Network (GNN) to model transfer-strategy labels as a function of schema-proficiency interactions. In this pilot implementation, each metaphor pair is represented as an instance with SCS, TCS, TPL, and the exploratory Sim_{schema} score as features. The graph links instances that share schema labels, comparable similarity ranges, or observed transfer-strategy categories, allowing the model to learn whether similar schema-proficiency configurations are associated with similar transfer strategies. Formally, the pilot graph is

$$G = (V, E, A), \quad V = \{v_i : (m_i, m'_i, x_i) \in D\}, \quad (7)$$

where each node v_i corresponds to one annotated metaphor pair and A is a weighted adjacency matrix. An edge weight is assigned when two nodes share at least one annotation feature:

$$A_{ij} = \lambda_1 \mathbb{I}[SCS_i = SCS_j] + \lambda_2 \mathbb{I}[TCS_i = TCS_j] + \lambda_3 \mathbb{I}[TPL_i = TPL_j] + \lambda_4 \text{sim}_{bin}(i, j). \quad (8)$$

Here $\mathbb{I}[\cdot]$ is an indicator function, $\lambda_1, \dots, \lambda_4$ are non-negative feature weights, and $\text{sim}_{bin}(i, j)$ equals 1 when two pairs fall into the same schema-similarity interval and 0 otherwise. The graph is undirected in the pilot implementation because shared annotation properties are treated symmetrically. This formulation does not assume that cultural schemas are ontologically fixed objects; it only defines an operational graph over adjudicated annotation labels.

For node v_i , the feature vector is

$$h_i^{(0)} = [\phi(SCS_i), \phi(TCS_i), \phi(TPL_i), \text{Sim}_{schema}(m_i, m'_i)]. \quad (9)$$

where $\phi(\cdot)$ denotes a one-hot or embedding representation of the corresponding categorical label. The GNN maps $h_i^{(0)}$ to a probability distribution over the four strategy labels in Eq. (3). The model minimizes the cross-entropy loss:

$$L(\theta) = - \sum_{i=1}^N \log P_{\theta}(T_i | SCS_i, TCS_i, TPL_i, \text{Sim}_{schema}(m_i, m'_i)) \quad (10)$$

The loss in Eq. (10) follows from maximum-likelihood estimation for a categorical label model: minimizing negative log likelihood increases the probability assigned to the adjudicated transfer-strategy labels. The model predicts transfer-strategy labels for held-out metaphor pairs by generalizing observed schema-proficiency patterns, such as advanced translators preferring Adaptation under moderate schema similarity. This prediction task evaluates label modeling within the annotation framework; it does not directly evaluate translation quality.

The computational cost is bounded by the size of the pilot graph. For a K -layer message-passing

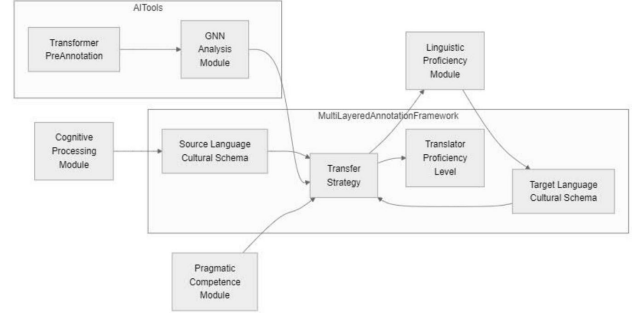


Fig. 1: Reformulated Cultural Schema Transfer Framework.

Source: created by the authors.

GNN with hidden dimension F , one forward pass has complexity $O(K|E|F)$ when sparse adjacency is used, plus $O(NF)$ for node-level classification. Because $N = 500$ in this pilot, the computational burden is modest. For larger corpora, edge construction should use sparse schema-neighborhood indexing rather than all-pairs comparison. Convergence in this study is empirical rather than theoretically guaranteed. The observed loss trend indicates stabilization after approximately 50 epochs, but this observation should be read as training behavior on the pilot corpus, not as a proof of convergence for arbitrary schema graphs.

4.4 Hybrid Human-in-the-Loop Annotation Pipeline

The annotation pipeline, shown in Figure 1, combines three stages. First, a BERT-wwm-assisted pre-annotation step generates candidate SCS and TCS labels using local textual context (e.g., “三纲五常” in a Confucian text $\rightarrow$ ConfucianMorality domain). Second, human annotators refine these candidate labels through active-learning-style prioritization, focusing first on low-confidence predictions and ambiguous metaphors such as “龙马精神.” Third, annotators verify transfer-strategy (T) and proficiency-level (TPL) tags, ensuring consistency with SCS/TCS alignments.

This pipeline addresses the scalability-reliability trade-off in corpus annotation, [9], achieving an inter-annotator agreement (Fleiss' κ) of 0.82 for SCS/TCS tags in pilot tests. Because the current study focuses on the final adjudicated annotations, the BERT component should be interpreted as an annotation-assistance tool rather than as an independently validated schema classifier.

Table 2 (Appendix) summarizes the computational components and their roles in the pilot workflow.

4.5 Integration with Translation Competence Models

The framework extends existing translation-competence models, [43], by treating cultural schema transfer as an annotatable and modelable component of translation competence. Research on adaptive, proficiency-aware AI systems suggests that computational support can be more useful when it responds to individual differences: adaptive frameworks responding to individual translator differences can enhance team performance in complex tasks, [44]; field evidence indicates that AI augmentation may improve performance, particularly for lower-proficiency users, [45]; and skill-aware AI systems may affect trust and effectiveness, [44], [45], [46]. Applied to schema-aware annotation, this evidence motivates proficiency-level feedback loops and adaptive recommendation strategies, although the present pilot does not directly test such interventions. The GNN's strategy predictions can be linked to a broader translation competence score that explicitly models proficiency moderation:

$$Competence_i = \alpha \cdot Sim_{schema}(m_i, m'_i) + \beta \cdot P_{\theta}(T_i | SCS_i, TCS_i, TPL_i, x_i) \quad (11)$$

where $\alpha, \beta \geq 0$ weight schema alignment and predicted strategy appropriateness. In the present pilot, this score is a descriptive construct rather than a validated assessment instrument. It can indicate where pedagogical feedback may be useful, for example when intermediate translators repeatedly choose Literal renderings for culturally distant metaphors, but future classroom studies would be needed before it could be used for learner evaluation.

By formalizing schema transfer as Eqs. (1)–(11), the framework provides a reproducible structure for annotation, reliability checking, and preliminary label prediction. The next section evaluates this structure through annotator training and pilot corpus analysis.

5 Annotator Training, Reliability Checks and Pilot Corpus Analysis

Implementing the multi-layered annotation scheme requires rigorous training protocols to ensure consistency in identifying cultural schemas and transfer strategies. This section outlines the annotator training process, inter-rater reliability assessments, and preliminary findings from a pilot corpus analysis.

The pilot corpus consists of 500 Chinese-English metaphor pairs and should be understood as a

purposive pilot sample rather than a representative population sample. Items were included when the source metaphor, target rendering, surrounding context, and translator-proficiency metadata were available for annotation. The corpus was designed to test whether the four-dimensional annotation scheme could be applied consistently and modeled preliminarily; it was not designed to establish domain-balanced estimates across all Chinese-English metaphor translation contexts. Domain balance and sampling bias are therefore treated as limitations for future corpus expansion.

5.1 Annotator Training Protocol

5.1.1 Training Phases

Training proceeds in three phases. The conceptual familiarization phase immerses annotators in cultural schema theory, focusing on Chinese-specific domains such as Confucianism and Daoism and their Western counterparts through contrastive analyses of metaphor pairs like “雪中送炭” (timely help) versus “a friend in need,” highlighting schema divergences. The strategy identification phase then equips trainees to classify transfer strategies using the taxonomy in Eq. (3), with practice exercises distinguishing literal translations such as “red-eye disease” for “红眼病” from adaptations such as “green-eyed monster.” The proficiency-level tagging phase requires annotators to review translator profiles, for example years of experience and training background, and assign TPL labels in Eq. (4), calibrated against standardized proficiency assessments, [47].

5.1.2 Calibration Sessions

To mitigate subjectivity, annotators participate in iterative calibration rounds where they independently tag the same metaphors and discuss discrepancies. For example, initial sessions revealed divergent interpretations of “龙马精神” (dragon-horse vigor), with some annotators tagging it under ConfucianMotivation and others under MythologicalEnergy. Consensus was reached by referring to corpus context, for example whether the text referenced classical literature or modern slogans.

5.2 Reliability and Validation

Inter-annotator agreement was measured using Fleiss' κ for categorical tags (SCS, TCS, TS) and intraclass correlation (ICC) for continuous schema similarity scores in Eq. (6). Results from the pilot corpus (500 Chinese-English metaphor pairs) showed reliable annotation behavior across the main dimensions. For SCS/TCS tags, $\kappa = 0.79$, with higher agreement for explicit schemas (e.g., ConfucianHierarchy: $\kappa = 0.85$) than implicit ones (e.g., DaoistNaturalism: $\kappa =$

0.72). For transfer strategies, $\kappa = 0.83$, with near-perfect agreement on Literal versus Substitution but lower consistency in distinguishing Adaptation from Omission. For schema similarity, ICC = 0.81 for CLIP-based scores, supporting their provisional use as an exploratory alignment feature rather than validating them as a complete measure of cultural equivalence.

Disagreements were resolved through adjudication by a senior linguist, refining the annotation guidelines, for example by adding “contextual dominance” rules to prioritize the most important schema when multiple interpretations exist.

5.3 Pilot Corpus Findings

The annotated pilot corpus revealed systematic patterns in how translators handle schema mismatches.

Proficiency-based differences were pronounced: intermediate translators defaulted to Literal translations (68% of cases) when $Sim_{schema} < 0.5$, often producing opaque outputs such as “three under five above” for “三纲五常,” whereas advanced translators employed Adaptation or Substitution (74% of cases) for the same metaphors, aligning schemas more effectively, for example “three moral bonds” with footnotes explaining Confucian context.

Schema similarity thresholds further shaped strategy distribution. Adaptation dominated when $0.5 \leq Sim_{schema} \leq 0.8$ (e.g., “画蛇添足” → “gilding the lily”), while Substitution occurred for $Sim_{schema} < 0.3$ (e.g., “红眼病” → “green-eyed monster”). Omission was rare (6%) and tied to extreme schema clashes ($Sim_{schema} < 0.2$), such as metaphors referencing Chinese historical events unknown in the target culture.

5.4 GNN Performance

The GNN model in Eq. (10) achieved 82% accuracy in predicting transfer-strategy labels on held-out data, outperforming a BERT-only label-prediction baseline (74%) by leveraging schema-proficiency interactions. For example, it correctly predicted that advanced translators would adapt “雪中送炭” to “a friend in need” ($Sim_{schema} = 0.65$), while intermediates rendered it literally as “delivering coal in snow.”

This result should be interpreted cautiously. The evaluation measures prediction of adjudicated annotation labels, not direct improvement in translation quality. The current pilot includes a BERT-only label-prediction comparison but does not include a full suite of machine-learning baselines. It also does not include human ratings of translation adequacy, fluency, or cultural appropriateness, nor does it include ablation tests isolating the

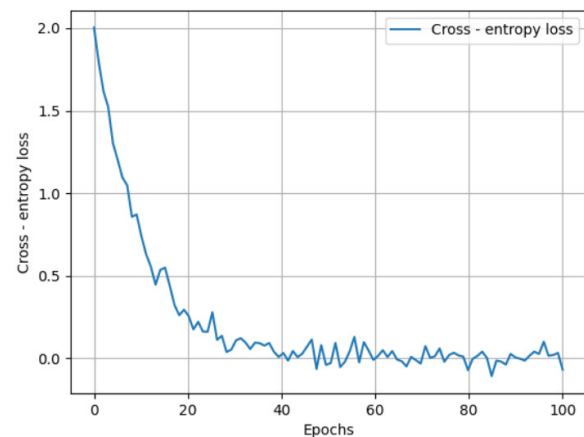


Fig. 2: Cross-entropy loss during GNN training.

Source: created by the authors.

contribution of each feature group. Table 3 (Appendix) summarizes the validation evidence available in the pilot study and the components reserved for future work. Repeated-run variance and external validation datasets are outside the scope of the present dataset. These components are therefore treated as priorities for future validation rather than as established evidence in this study.

To make the statistical status of the pilot results explicit, the 82% label-prediction accuracy can be read as a descriptive estimate over the available annotated corpus. If the 500 metaphor pairs are used as the denominator, a binomial Wilson interval gives an approximate 95% confidence interval of 78.4%–85.1%. This interval is reported only as an uncertainty estimate for the observed pilot accuracy; it should not be interpreted as evidence of population-level generalization, because the corpus is purposive and not a random sample of Chinese-English metaphor translations.

Figure 2 shows the GNN’s learning curve, with loss plateauing after 50 epochs, indicating stable strategy-label prediction. Table 4 (Appendix) reports the available training metrics across epochs. Cross-entropy loss decreases steadily, while validation accuracy plateaus at approximately 82% after 50 epochs; early stopping criteria were applied after epoch 60 to reduce overfitting risk.

Figure 3 illustrates the positive correlation between Sim_{schema} and prediction accuracy ($r = 0.78$), suggesting that schema alignment is associated with transfer-strategy predictability in this pilot corpus. Table 5 (Appendix) reports accuracy metrics across schema similarity ranges for the 500 metaphor pairs in the pilot corpus. Higher schema similarity (0.9–1.0) and very low schema similarity (0.0–0.2) are associated with more predictable

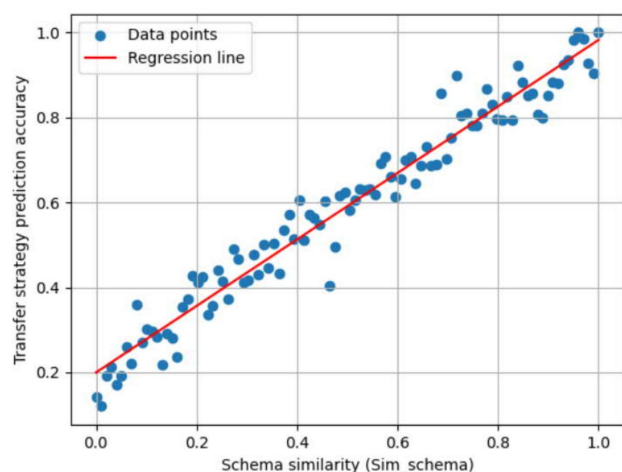


Fig. 3: Schema similarity vs. transfer strategy accuracy.

Source: created by the authors.

transfer strategies, whereas intermediate ranges (0.3–0.7) show greater variability. The relationship should not be interpreted as evidence that schema similarity alone determines translatability.

5.5 Limitations and Adjustments

Three notable limitations emerged from the pilot study. First, metaphors tied to low-frequency schemas, for example Buddhist karma, had sparse data, reducing GNN accuracy; future work will need larger corpora or carefully controlled augmentation to improve coverage of underrepresented cultural domains. Second, contextual ambiguity posed challenges, as some metaphors such as “龙马精神” required paragraph-level context for schema disambiguation, prompting guideline updates to mandate context windows. Third, the 500-pair dataset is sufficient for a pilot annotation study but small for drawing strong claims about GNN generalization. These limitations notwithstanding, the pilot results support the feasibility of the four-dimensional annotation scheme and indicate that schema-strategy-proficiency interactions can be modeled preliminarily within a Chinese-English corpus.

6 Discussion and Future Work

6.1 Limitations of the Proposed Method

The multi-layered annotation framework remains limited in several respects. Pre-trained models like CLIP may inadvertently introduce biases when used for exploratory schema similarity. Research converges on the pervasiveness of cultural biases in multimodal embeddings: gender bias has been documented in vision-language models,

[34], and generative AI models exhibit systematic cultural tendencies based on training language composition, [48]. Chinese-language metaphors may therefore be conflated with Western conceptual domains due to imbalanced multilingual training. The present study does not include a dedicated empirical bias audit, so CLIP-based scores should be treated as heuristic signals subject to human adjudication. Proficiency-level classification (intermediate/advanced) necessarily oversimplifies translator competence, which often varies across domains and metaphor types. The GNN’s strategy predictions, though accurate for high-frequency schemas, degrade when handling rare or hybrid metaphors, for example “佛系” or “Buddha-like” indifference, as noted in Section 5.4. Finally, because the pilot corpus is limited to Chinese-English metaphor pairs, the findings should not be generalized to other language pairs without further validation.

6.2 Potential Application Scenarios

The framework’s granular annotations have possible uses in both pedagogy and technology. In translator training, schema-strategy mappings could support adaptive learning systems that recommend exercises based on a learner’s proficiency and the detected schema mismatch. For example, an intermediate translator repeatedly opting for literal renditions of Confucian metaphors might receive targeted exercises on adaptation techniques. Such applications remain prospective because the present study does not evaluate learner outcomes.

In NLP, the annotated corpus could inform future neural machine translation systems by providing schema-aware labels that help diagnose metaphor-related errors. The present study does not test downstream NMT quality, so claims about translation improvement remain prospective. By integrating human expertise with automated processing, human-in-the-loop approaches could address important limitations in handling metaphor-rich and culturally complex texts, provided that future work evaluates actual translation outputs through baseline comparisons and human quality ratings.

Cross-modal applications also warrant later study, including culturally coherent image captions or subtitles that align visual metaphors, for example dragon motifs, with target-language schemas. However, such systems must account for cultural biases in pre-trained embeddings and vision-language models. Careful curation of schema annotations to represent diverse cultural perspectives is necessary.

6.3 Comparison with Other Metaphor Translation Methods

The framework's novelty lies in its explicit modeling of schema-strategy-proficiency interactions, which prior methods address only partially. Conventional transfer-based approaches, [33], focus on lexical or syntactic equivalences without schema tagging, while cognitive models, [3], prioritize conceptual mappings but neglect translator variability. Hybrid methods, [2], annotate strategies and schemas separately, missing their interdependencies, for example how Adaptation correlates with high schema similarity. By contrast, the GNN-based approach models these relationships as a strategy-label prediction task, achieving 82% accuracy in the pilot corpus (Section 5.4). However, this result should be interpreted as preliminary annotation-modeling evidence rather than as proof of superior translation performance. The computational overhead of GNNs may also limit real-time applications, suggesting a trade-off between interpretability and scalability that future work could address via lightweight graph architectures, [1].

6.4 Future Work

Four directions follow from these limitations. Expanding the schema inventory to include diachronic variations, for example how “龙马精神” shifts from mythological to modern motivational schemas, could better capture cultural evolution in metaphors and examine how cultural context shifts influence translator strategy selection across proficiency levels. Integrating multimodal data, for example images or gestures associated with metaphors, might enrich schema representations, particularly for embodied metaphors like “心花怒放” (heart flowers bursting with joy), but such work should include explicit bias checks for multimodal embeddings. Longitudinal studies tracking translators' strategy use over time could test whether schema-aware annotation has pedagogical effects. Future computational evaluation should include stronger baselines, ablation studies, repeated-run variance, statistical significance testing, and human evaluation of translation quality.

7 Conclusion

The multi-layered cultural schema annotation framework presented in this study provides a systematic pilot approach to analyzing metaphor translation between Chinese and English. By dissecting the translation process into four interdependent dimensions—source and target cultural schemas, transfer strategies, and translator proficiency levels—the framework provides a

structured methodology for making cultural-schema decisions explicit and comparable. The integration of BERT-wwm-assisted pre-annotation, exploratory CLIP-based similarity scoring, human adjudication, and Graph Neural Network label prediction demonstrates how selected aspects of cultural schema transfer can be represented as annotatable and empirically modelable relationships.

The pilot corpus analysis provides preliminary support for the framework's feasibility, revealing patterns in how translators at different proficiency levels navigate schema mismatches. Advanced translators more often employ adaptation and substitution strategies when faced with culturally distant metaphors, while intermediate learners tend to default to literal translations. These findings suggest the pedagogical value of explicit schema annotation in translator training, where awareness of cultural-conceptual mappings may support competence development. For computational applications, the framework's annotated data provide culturally grounded training signals for future work, though downstream translation-quality improvement remains to be tested.

Looking ahead, the framework's modular design allows for expansion into new cultural domains and multimodal contexts. Future iterations could incorporate diachronic schema variations or visual-gestural metaphors, further enriching the analysis of cross-cultural communication. The current limitations, such as biases in pre-trained embeddings, the oversimplification of proficiency levels, the small pilot dataset, and the absence of direct translation-quality evaluation, present opportunities for refinement through more diverse training data and finer-grained competence metrics. Ultimately, this work contributes a cautious pilot model for transforming cultural awareness from an implicit skill into an annotatable and teachable component of translator expertise. Broader implications for other language pairs remain promising but require independent empirical validation.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

The authors wrote, reviewed and edited the content as needed and they have not utilised artificial intelligence (AI) tools. The authors take full responsibility for the content of the publication.

References:

- [1] A. Chowdhery, S. Narang, J. Devlin, M. Bosma, G. Mishra, A. Roberts, P. Barham, H. W. Chung, C. Sutton, S. Gehrmann, P. Schuh, K. Shi, S. Tsvyashchenko, J. Maynez, A. Rao, P. Barnes, Y. Tay, N. Shazeer, V. Prabhakaran, E. Reif, N. Du, B. Hutchinson, R. Pope, J. Bradbury, J. Austin, M. Isard, G. Gur-Ari, P. Yin, T. Duke, A. Levskaya, S. Ghemawat, S. Dev, H. Michalewski, X. Garcia, V. Misra, K. Robinson, L. Fedus, D. Zhou, D. Ippolito, D. Luan, H. Lim, B. Zoph, A. Spiridonov, R. Sepassi, D. Dohan, S. Agrawal, M. Omernick, A. M. Dai, T. S. Pillai, M. Pellat, A. Lewkowycz, E. Moreira, R. Child, O. Polozov, K. Lee, Z. Zhou, X. Wang, B. Saeta, M. Diaz, O. Firat, M. Catasta, J. Wei, K. Meier-Hellstern, D. Eck, J. Dean, S. Petrov, and N. Fiedel, "PaLM: Scaling language modeling with pathways," *Journal of Machine Learning Research*, vol. 24, no. 240, pp. 1–113, 2023. <https://jmlr.org/papers/v24/22-1144.html> (accessed September 4, 2026).
- [2] H. Zhu, O. H. A. Mahfoodh, and D. T. A. Lin, "A study of corpus construction for English metaphorical translation using natural language processing," *Pakistan Journal of Life & Social Sciences*, vol. 22, no. 1, pp. 5629–5639, 2024. https://www.pjlss.edu.pk/pdf_files/2024_1/5629-5639.pdf (accessed September 4, 2026).
- [3] S. He, "Cognitive metaphor theories in translation studies: Toward a dual-model parametric approach," *Intercultural Pragmatics*, vol. 18, no. 1, pp. 25–52, 2021. <https://doi.org/10.1515/ip-2021-0002>.
- [4] Z. Kövecses and G. Radden, "Metonymy: Developing a cognitive linguistic view," *Cognitive Linguistics*, vol. 9, no. 1, pp. 37–78, 1998. <https://doi.org/10.1515/cogl.1998.9.1.37>.
- [5] G. Lakoff, "The contemporary theory of metaphor," in *Metaphor and Thought*, A. Ortony, Ed. Cambridge University Press, 1993, pp. 202–251. <https://doi.org/10.1017/CB09781139173865.013>.
- [6] L. Hui, "Chinese cultural schema of education: Implications for communication between Chinese students and Australian educators," *Issues in Educational Research*, vol. 15, no. 1, pp. 17–36, 2005. <https://search.informit.org/doi/10.3316/informit.051194588652141> (accessed September 4, 2026).
- [7] Z. Li, "A comparative survey of vision metaphors based on the corpus in English and Chinese," *Theory and Practice in Language Studies*, vol. 3, no. 7, pp. 1232–1242, 2013. <https://doi.org/10.4304/tp1s.3.7.1232-1242>.
- [8] Z. Xu and N. Fang, "Re-schematisation of Chinese xiao (filial piety) across cultures and generations," in *Cultural Linguistics and World Englishes*, M. Sadeghpour and F. Sharifian, Eds. Springer, 2021, pp. 65–85. https://doi.org/10.1007/978-981-15-4696-9_4.
- [9] J. Newman and C. Cox, "Corpus annotation," in *A Practical Handbook of Corpus Linguistics*, M. Paquot and S. Th. Gries, Eds. Springer, 2020, pp. 25–48. https://doi.org/10.1007/978-3-030-46216-1_2.
- [10] W. Teubert, "The role of parallel corpora in translation and multilingual lexicography," in *Lexis in Contrast: Corpus-based Approaches*, B. Altenberg and S. Granger, Eds. John Benjamins, 2002, pp. 189–214. <https://doi.org/10.1075/sc1.7.14teu>.
- [11] L. Bentivogli and E. Pianta, "Exploiting parallel texts in the creation of multilingual semantically annotated resources: The MultiSemCor Corpus," *Natural Language Engineering*, vol. 11, no. 3, pp. 247–261, 2005. <https://doi.org/10.1017/S1351324905003839>.
- [12] E. Sanchez-Bayona and R. Agerri, "Meta4XNLI: A cross-lingual parallel corpus for metaphor detection and interpretation," *Computational Linguistics*, vol. 52, no. 1, pp. 191–235, 2026. <https://doi.org/10.1162/coli.a.20>.
- [13] S. A. M. Asiri and A. A. Metwally, "The impact of linguistic and cultural competence on translation quality: Pedagogical insights into translation problems," *Journal of Language Teaching and Research*, vol. 11, no. 3, pp. 509–520, 2020. <https://doi.org/10.17507/jltr.1103.22>.
- [14] G. M. Shreve, "Professional translator development from an expertise perspective," in *The Bloomsbury Companion to Language Industry Studies*, E. Angelone, G. Massey, and M. Ehrensberger-Dow, Eds. Bloomsbury Publishing, 2019, pp. 153–178. <https://doi.org/10.5040/9781350024960.0012>.

- [15] E. Shutova, B. J. Devereux, and A. Korhonen, "Conceptual metaphor theory meets the data: A corpus-based human annotation study," *Language Resources and Evaluation*, vol. 47, no. 4, pp. 1261–1284, 2013. <https://doi.org/10.1007/s10579-013-9238-z>.
- [16] K. Hu, *Introducing Corpus-based Translation Studies*. Springer, 2016. <https://doi.org/10.1007/978-3-662-48218-6>.
- [17] F. Dalayli, "Use of NLP techniques in translation by ChatGPT: Case study," in *Proceedings of the Workshop on Computational Terminology in NLP and Translation Studies (ConTeNTS) Incorporating the 16th Workshop on Building and Using Comparable Corpora (BUCC)*, INCOMA Ltd., 2023, pp. 19–25. <https://aclanthology.org/2023.contents-1.3>.
- [18] A. Haddad Haddad, "Image schemas and image schematic complexes: Enhancing neural machine translation networks," in *Computational and Corpus-Based Phraseology*, G. Corpas Pastor and R. Mitkov, Eds. Springer, 2022, pp. 105–115. https://doi.org/10.1007/978-3-031-15925-1_8.
- [19] C. Schäffner, "Metaphor and translation: Some implications of a cognitive approach," *Journal of Pragmatics*, vol. 36, no. 7, pp. 1253–1269, 2004. <https://doi.org/10.1016/j.pragma.2003.10.012>.
- [20] J. Wang, "Translating and rewriting Chinese proverbs: A case study of Howard Goldblatt's English translation of Mo Yan's *Shengsi Pilao*," *English Language and Literature Studies*, vol. 9, no. 2, pp. 12–24, 2019. <https://doi.org/10.5539/ells.v9n2p12>.
- [21] R. W. Gibbs Jr., "Evaluating conceptual metaphor theory," *Discourse Processes*, vol. 48, no. 8, pp. 529–562, 2011. <https://doi.org/10.1080/0163853X.2011.606103>.
- [22] N. Yu, "Chinese metaphors of thinking," *Cognitive Linguistics*, vol. 14, no. 2–3, pp. 141–165, 2003. <https://doi.org/10.1515/cogl.2003.006>.
- [23] M. Farghal, "Literary translation: A schema-theoretic model," *Al-'Arabiyya*, vol. 37, pp. 21–35, 2004. <http://www.jstor.org/stable/43195529> (accessed September 4, 2026).
- [24] P. Newmark, *A Textbook of Translation*. Prentice Hall, 1988. <https://archive.org/details/textbookoftransl0000newm> (accessed September 4, 2026).
- [25] G. Toury, *Descriptive Translation Studies and Beyond*. John Benjamins, 1995. <https://doi.org/10.1075/btl.4>.
- [26] L. Venuti, *The Translator's Invisibility: A History of Translation*. Routledge, 1995. <https://doi.org/10.4324/9780203360064>.
- [27] M. Tymoczko, "Western metaphorical discourses implicit in translation studies," in *Thinking Through Translation with Metaphors*, J. St. Andre, Ed. Routledge, 2010. <https://doi.org/10.4324/9781315759746-6>.
- [28] C. Schäffner, "Political discourse analysis from the point of view of translation studies," *Journal of Language and Politics*, vol. 3, no. 1, pp. 117–150, 2004. <https://doi.org/10.1075/jlp.3.1.09sch>.
- [29] X. Kong, H. Fang, W. Chen, J. Xiao, and M. Zhang, "Examining human-AI collaboration in hybrid intelligence learning environments: Insight from the Synergy Degree Model," *Humanities and Social Sciences Communications*, vol. 12, no. 1, p. 821, 2025. <https://doi.org/10.1057/s41599-025-05097-z>.
- [30] M. Steyvers, H. Tejada, G. Kerrigan, and P. Smyth, "Bayesian modeling of human-AI complementarity," *Proceedings of the National Academy of Sciences*, vol. 119, no. 11, e2111547119, 2022. <https://doi.org/10.1073/pnas.2111547119>.
- [31] A. Nguyen, Y. Hong, B. Dang, and X. Huang, "Human-AI collaboration patterns in AI-assisted academic writing," *Studies in Higher Education*, vol. 49, no. 5, pp. 847–864, 2024. <https://doi.org/10.1080/03075079.2024.2323593>.
- [32] G. Stockwell and Y. Wang, "Framing human-AI collaboration in language education," *Technology in Language Teaching & Learning*, vol. 7, no. 4, 103824, 2025. <https://doi.org/10.29140/tl.v7n4.103824>.
- [33] E. S. Fernández, "Translation studies and the cognitive theory of metaphor," *Review of Cognitive Linguistics*, vol. 9, no. 1, pp. 262–279, 2011. <https://doi.org/10.1075/rcl.9.1.12sam>.

- [34] M. Bartl, A. Mandal, S. Leavy, and S. Little, "Gender bias in natural language processing and computer vision: A comparative survey," *ACM Computing Surveys*, vol. 57, no. 6, pp. 1–36, 2025. <https://doi.org/10.1145/3700438>.
- [35] J. Moorkens, S. Castilho, F. Gaspari, and S. Doherty, *Translation Quality Assessment: From Principles to Practice*. Springer, 2018. <https://doi.org/10.1007/978-3-319-91241-7>.
- [36] S. Zhang, Y. Zhuang, and L. Chang, "Cultural mediation in crisis translation: A snapshot of the citizen translator in China's Greater Bay Area during the COVID-19 pandemic," *Translation and Interpreting Studies*, vol. 18, no. 2, pp. 301–324, 2023. <https://doi.org/10.1075/tis.23010.zha>.
- [37] S. Bruti, "Cross-cultural pragmatics: The translation of implicit compliments in subtitles," *The Journal of Specialised Translation*, no. 6, pp. 185–197, 2006. <https://doi.org/10.26034/cm.jostrans.2006.747>.
- [38] F. Zanettin, "Parallel corpora in translation studies: Issues in corpus design and analysis," in *Intercultural Faultlines*, M. Olohan, Ed. Routledge, 2017, pp. 105–118. <https://doi.org/10.4324/9781315759951-8>.
- [39] N. Duran, S. Battle, and J. Smith, "Inter-annotator agreement using the conversation analysis modelling schema, for dialogue," *Communication Methods and Measures*, vol. 16, no. 3, pp. 182–214, 2022. <https://doi.org/10.1080/19312458.2021.2020229>.
- [40] M. A. Taylor and D. S. Stoltz, "A workflow for analyzing cultural schemas in texts," *The Journal of Mathematical Sociology*, vol. 49, no. 1, pp. 1–24, 2025. <https://doi.org/10.1080/0022250x.2024.2395806>.
- [41] P. Atzeni, P. Cappellari, R. Torlone, P. A. Bernstein, and G. Gianforme, "Model-independent schema translation," *The VLDB Journal*, vol. 17, no. 6, pp. 1347–1370, 2008. <https://doi.org/10.1007/s00778-008-0105-2>.
- [42] A. Barcelona, "On the systematic contrastive analysis of conceptual metaphors: Case studies and proposed methodology," in *Applied Cognitive Linguistics: Language Pedagogy*, M. Pütz, S. Niemeier, and R. Dirven, Eds. Walter de Gruyter, 2001, pp. 117–146. <https://doi.org/10.1515/9783110866254.117>.
- [43] M. Hanne, "Metaphors for the translator," in *The Translator as Writer*, S. Bassnett and P. Bush, Eds. Bloomsbury Publishing, 2007, pp. 208–224. <https://books.google.com/books?hl=zh-TW&lr=&id=lvjeBAAAQBAJ> (accessed September 4, 2026).
- [44] M. Zhao, R. Simmons, and H. Admoni, "The role of adaptation in collective human–AI teaming," *Topics in Cognitive Science*, vol. 17, no. 2, pp. 291–323, 2025. <https://doi.org/10.1111/tops.12633>.
- [45] C. Liao, X. Wen, S. Li, and P. Du, "How effective is AI augmentation in human–AI collaboration? Evidence from a field experiment," *Information Technology & People*, vol. 37, no. 7, pp. 2357–2389, 2024. <https://doi.org/10.1108/ITP-11-2022-0859>.
- [46] T. Du, X. Li, N. Jiang, Y. Xu, and Y. Zhou, "Adaptive AI as collaborator: Examining the impact of an AI's adaptability and social role on individual professional efficacy and credit attribution in human–AI collaboration," *International Journal of Human–Computer Interaction*, vol. 41, pp. 12422–12433, 2025. <https://doi.org/10.1080/10447318.2025.2462120>.
- [47] V. Strilets, "Approaches to assessing translation proficiency," *International Journal of Philology*, vol. 27, no. 4, pp. 89–98, 2023. [https://doi.org/10.31548/philolog14\(4\).2023.09](https://doi.org/10.31548/philolog14(4).2023.09).
- [48] J. G. Lu, L. L. Song, and L. D. Zhang, "Cultural tendencies in generative AI," *Nature Human Behaviour*, vol. 9, pp. 2360–2369, 2025. <https://doi.org/10.1038/s41562-025-02242-1>.

APPENDIX

Table 1. Notation for the pilot annotation and label-prediction framework.

Symbol	Definition
m_i	Chinese source-language metaphor in the i th aligned pair
m'_i	English target-language rendering of m_i
x_i	Local textual context used by annotators and pre-annotation tools
SCS_i	Source-language cultural-schema annotation for m_i
TCS_i	Target-language cultural-schema annotation for m'_i
$d_{s,i}, d_{t,i}$	Source and target cultural domains
$s_{s,i}, s_{t,i}$	Source and target schema labels within the corresponding domains
T_i	Observed transfer-strategy label after adjudication
TPL_i	Translator proficiency label associated with the aligned pair
Sim_{schema}	Exploratory CLIP-based similarity score between source and target schema descriptions
θ	Trainable parameters of the probabilistic label-prediction model

Source: created by the authors.

Table 2. Computational components of the pilot annotation workflow.

Component	Role in the framework	Inputs	Outputs	Validation in this study
BERT-wwm-assisted pre-annotation	Candidate schema-label generation	Metaphor pair and local context	Candidate SCS/TCS labels	Human correction and adjudication; not evaluated as a standalone classifier
CLIP-based similarity feature	Exploratory schema-alignment score	Short source/target schema descriptions	Sim_{schema} score	ICC for similarity-score consistency; interpreted as heuristic
Human adjudication	Final annotation authority	Candidate labels and annotation guidelines	Final SCS/TCS/TS/TPL labels	Fleiss' κ for categorical tags
GNN strategy-label model	Prediction of observed transfer-strategy labels	SCS, TCS, TPL, and Sim_{schema}	Predicted TS label	Held-out label accuracy against adjudicated annotations

Source: created by the authors.

Table 3. Pilot validation evidence and future validation work.

Validation component	Status in the present study	Interpretive consequence
Held-out label prediction	Reported pilot accuracy of 82%	Supports feasibility of modeling adjudicated strategy labels
BERT-only comparison	Reported baseline accuracy of 74%	Provides an initial comparison, but not a comprehensive benchmark suite
Uncertainty estimate	Approximate Wilson interval of 78.4%–85.1% if $N = 500$ is treated as the descriptive denominator	Indicates sampling uncertainty around the pilot accuracy estimate
Cross-validation and repeated random seeds	Not available in the current dataset	Required before making stability claims
Ablation study	Not available in the current dataset	Required to isolate the contribution of SCS, TCS, TPL, and Sim_{schema}
External validation corpus	Not available in the current dataset	Required before claiming generalization beyond the pilot corpus
Human translation-quality ratings	Not collected in the current study	Required before claiming downstream translation-quality improvement

Source: created by the authors.

Table 4. GNN training performance metrics.

Epoch	Cross-Entropy Loss	Validation Accuracy	Learning Rate
10	0.847	71.2%	0.001
20	0.612	76.5%	0.001
30	0.389	79.8%	0.0005
40	0.256	81.1%	0.0005
50	0.143	82.0%	0.0005
60	0.141	82.1%	0.00025
70	0.140	82.2%	0.00025

Source: created by the authors.

Table 5. Transfer strategy distribution by schema similarity range.

Schema Range	Similarity	Primary Strategy	Transfer	Secondary Strategy	Prediction Accuracy	Sample Size	Examples
0.0–0.2		Omission		Substitution	84.3%	39	红眼病 → (omitted)
0.2–0.3		Substitution		Omission	79.1%	74	红眼病 → green-eyed monster
0.3–0.5		Substitution		Adaptation	76.5%	112	龙马精神 → dragon-like vigor
0.5–0.7		Adaptation		Literal	81.2%	130	画蛇添足 → gilding the lily
0.7–0.9		Literal		Adaptation	77.8%	82	三人行 → three who walk together
0.9–1.0		Literal		Adaptation	85.6%	63	友谊 → friendship

Source: created by the authors.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

Dahui Dong was responsible for the study design, corpus annotation framework, data interpretation, manuscript writing, and revision.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflicts of Interest

The author has no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0:
https://creativecommons.org/licenses/by/4.0/deed.en_US